

Has the peace dividend vanished?

Hopes that the end of the Cold War would revitalize basic research have been disappointed for good reasons, but politicians must restrain their expectations of research — and their impatience.

WHATEVER happened to the “peace dividend”? When the Berlin Wall came down at the end of 1989, there were eager calculations of how much extra could then be spent on good causes, basic research conspicuously. But that has not happened. Indeed, grant-making agencies are now as short of funds as in the 1980s, with the result that good projects go begging. In the former Soviet Union, the needs of research have been overshadowed by those of survival. And in the United States, even talismanic projects such as the Superconducting Super Collider (SSC) have been scrapped. What is happening to the research community, why, and what will be the consequences?

The discovery that research is not the first beneficiary of a relatively peaceful world is not, of course, new. That, as it happens, was the theme of *Nature's* news survey at the beginning of this year. The evidence that the end of the Cold War does not mean extra for basic research is clear from how governments throughout the West have cut back on the funds that granting agencies have to spend. But the causes of this state of affairs are obscure. And it is unlikely that it will be possible to put things right until they are understood.

One remarkable feature of the present wariness of basic research is that governments appear to have stumbled on it independently, without collusion in advance. (The decision of the US Congress last November to cancel the SSC may encourage other governments to draconian measures of their own, but that is still a threat for the future.) The common influence on them is economic. In retrospect, it is not surprising that the ending of the Cold War should have had a deflationary influence. Part of what governments used to spend on military affairs is instead consumed by welfare payments to those put out of work. But Western governments are especially afflicted by the way in which funds that would normally be invested in their own territories are instead being spent elsewhere (see *Nature* 368, 379; 1994). So the public purse is everywhere straitened, while the demands on it have multiplied. Governments can easily claim that there is no more money for research.

Accountability

But more than that has happened. In the United States, Senator Barbara Mikulski (Democrat, Maryland) is only the most forceful (or widely reported) of those who have been insisting that the research enterprise does not exist for its own sake, but for the benefits that its work will bring the wider community. In Britain, the government has more openly

taken the view that the civil research establishment (costing something like £1 billion a year) must be harnessed to the cause of “wealth creation”. (Only this month will there be evidence of what that means in the practice of making research grants.) Governments, in other words, are saying that their part of the research enterprise must not merely earn its own keep, but also add to national prosperity.

That is where the argument begins to go awry. Nobody would ask that the research enterprise should be kept in the condition to which it has become used for decorative purposes only. In any case, part of the pride of science is that its activities over the past two centuries, has created prosperity previously unimaginable. But even agreeing that those who spend public money should be accountable, demands that public research budgets should be required as extra national prosperity in the here and now are misplaced.

Benefits

Unrealistic? The familiar experience of the past two centuries is that there is commonly a huge time-lag between a discovery in basic science and some kind of prosperity wrung from it. Take the silicon chip, for example: should the credit go the US Department of Defense for having sponsored the development of densely packed integrated transistor circuits in the early 1970s, to Shockley, Bardeen and Brattain for having invented the transistor in 1949 or to Nevill Mott for having explained in the 1930s how metal-oxide junctions would function before anybody knew what use to put them to? Or the laser, now a part of every country's gross national product: does that belong to Townes, Basov and Prokhorov (an American and two Russians) or to Einstein (still then a Swiss) for his discovery (or, more accurately, prediction) of stimulated emission? If politicians are not disabused of the belief that research can yield quick benefits, it will not be long before they are disillusioned and begin proclaiming that it is altogether a waste of money.

And misplaced? The general opinion that basic science is an international enterprise cannot easily be reconciled with governments' wish that it should enhance national prosperity. So much is vividly illustrated by the authors' list of the account of the total synthesis of taxol published the other week (*Nature* 367, 630–634; 1994); Ralph A. Lewin of the Scripps Institute of Oceanography has remarked that the surnames of the 10 authors betoken origins in 8 different countries. There is no obvious way in which the benefits of this remarkable feat can be equitably reassigned to the



several national economies concerned. Nor is one necessary.

The truth is that the impatience and chauvinism that mark governments' expectations in the wake of the Cold War are enemies of basic research. But the expectations are also misplaced because they discount the value of the most direct and immediate of its benefits for modern economies — the present value of the skilled people who emerge from a spell in basic research who are uniquely equipped to change the world in more practical ways. There are all kinds of unmade calculations to be made about the cost per head of training people in ways such as this; the outcome would amply justify what is spent on research. That is what governments must understand about the Cold War and the future. □

Freedom's end?

The US Congress should find out whether NASA has a real design for its space station — and then cancel it.

WITH the unfolding of the budget cycle in the US Congress, people have begun to suck their teeth again about the plan to build a space station, which survived last year's cycle only by a single vote in the House of Representatives. This year, it must be hoped, the Congress will be more resolute. Its best course would be to cancel the project. Otherwise, it must take steps to limit the damage its survival will cause to the more valuable parts of the US space programme.

At this stage, very little is required to establish that simple truth; the few benefits the space station would bring are not justified by its expected cost, while it cannot safely be afforded within the budget of the National Aeronautics and Space Administration (NASA) that the administration (with an eye on the federal deficit) has decreed acceptable between now and the end of the decade. Worse than that, far from being the technically superb counterpoise to the ex-Soviet space-station *Mir* planned a decade ago, space station *Freedom* has all the makings of a technical bungle on a grand scale. During the months ahead, the Congress should be particularly careful to enquire about the last point.

The chequered history of the project points clearly to the hazards, but does not reveal what *Freedom* will actually be like. To be sure, when almost anybody with access to a PC and a decent graphics program can run up a conceptual design, much has been said to suggest what it might be like: invariably, there is a kind of tank for housing people, docking stations at which visitors would be received and extended booms on which solar panels and other sensitive equipment would be mounted. But it is a far cry from such a specification to a design that has been tried and tested.

NASA itself is only partly to blame for this state of affairs. It has been required on several occasions in the past decade to cut its coat to suit the diminished cloth available. The most taxing of those occasions was last year when NASA was required to produce, within ten days, a range of three "cheaper" designs so that President Bill Clinton, seeking to avoid a budgetary showdown, could choose one of them. By all accounts, the computer-buffs were able marvellously to

exercise their graphics programs. What the congressional committees should insist on knowing is whether the design now chosen is indeed a design in the engineering sense. When those who make motorcars dare not put on the market products whose components have not been tested, NASA should be given firmly to understand that it cannot risk people's lives to equipment whose design is conceptual only.

There is also a political issue to be discounted. Last year, NASA agreed to collaborate with the ex-Soviet space programme in a variety of vaguely specified projects whose overall purpose is the training of those who will eventually occupy and operate space-station *Freedom*. Helping Russia is politically correct, so much so (in NASA's view) that the Congress will not dare demur. But to what end? To keep the now-Russian programme in being to serve as a potential competitor for prestige (and consequently higher budgets) at some stage in the future? Or does NASA reckon that it may yet get space-station *Freedom* by buying another *Mir* off the Russian shelf? What exactly does NASA have in mind?

The objective case for *Freedom* is something else again, and is insupportable. To be sure, it would be mildly interesting to have a stream of data about phenomena such as the crystallization of solids in nearly zero gravity, but nobody would pay for them at the expected cost, \$8 billion for the first set of hardware and several billion dollars a year for operations. But all that is established. The real worry is that NASA's budget ceiling of roughly \$14.5 billion a year (the arguments in the next few months will be about the odd \$100 million) is already tight, and will be tighter still when *Freedom*'s declared cost begins to escalate. Then the squeeze will be transferred to NASA's exploration of the Solar System, 'Mission to Planet Earth' included. That would be a discreditable resolution. Better that the Congress should draw a ring-fence around the serious part of NASA's work (aeronautics as well as planetary science), saying that *Freedom* might be paid for out of the small change left over. □

Less advice is just fine

US budget office kills 284 federal advisory committees.

THAT which is created by government usually lasts forever, often long past its time. Therefore, word that the US Office of Management and Budget (OMB) has terminated 284 government advisory committees (and thereby saved an estimated \$17 million) comes as a breath of fresh air. Advice from duly chosen experts is, of course, a valuable thing. But there comes a point when the cost of advice exceeds its value.

Does the United States really need a Network Reliability Council, or a Technical Advisory Group for Cigarette Fire Safety? And, much as we care about pigs, is an Advisory Committee on Swine Health Protection really in their best interest? The OMB has judiciously said "*finis*" to these and other advisory groups including, one whose duty is to "Take Pride in America", and one that, since 1897, has been solemnly charged with sniffing imported tea. OMB should stoutly resist efforts to reincarnate them. □

Holocaust denial research disclaimed

Munich. The Max Planck Society has reached an out of court settlement with a former graduate student whom it sacked last summer for producing research which "proved" that gassing of prisoners at Auschwitz concentration camp never happened.

The student, Germar Rudolf, was fired without notice from the Max Planck Institute of Solid State Physics in Stuttgart last June, after senior staff members found out that he had used his position at the institute to carry out the research, the results of which have been widely circulated by groups who deny that the holocaust took place.

The affair has left the Max Planck Society, which runs dozens of prestigious research institutes across Germany, angry that its name has been tarred by association with the holocaust denial movement, and Rudolf bitter that he gets no compensation and has little chance of finding another job. The March 22 settlement pays Rudolf no compensation, but rewords his dismissal as an "ending of the contract through mutual agreement".

The story began several years ago when ex-Nazi general Otto Ernst Remer — who led the suppression of the *putsch* against Hitler in July 1944 — was charged with inciting race hatred. He had long campaigned for an end to what he calls the "Auschwitz lies", claiming that the holocaust had not happened.

During his trial, his lawyer Hajo Herrmann commissioned Rudolf to conduct experiments which could be used to support Remer's claim. Contact between the two had been established after Rudolf had a letter published in a newspaper about the Leuchter report, which had made claims that mass exterminations could not have happened at Auschwitz.

At the time Rudolf had just started his PhD at the Institute for Solid State Physics, where he was initially considered to be a good scientist, earnestly concerned about Germany's past, but not necessarily a fascist.

He travelled to Auschwitz in summer 1991, and took samples from the walls of one of the gas chambers and one of the delousing chambers. Using Max Planck stationery, he sent the samples to an analytical laboratory in Taunusstein, the Fresenius Institute.

Analysis showed no detectable cyanide in the samples taken from the gas chamber, but detectable cyanide in samples taken from the delousing chamber. These results formed the basis of a 120-page report in which he elaborated his theory — also based on an analysis of the structure of the gas chambers — that the gassings could

not have taken place.

Whether Rudolf was paid by Remer's lawyer for his work is not clear, although he was certainly given expenses. Rudolf had presented his report to the lawyer on the understanding that it would be used only in the court case and would not be distributed more widely. As it turned out, the report was not used as evidence in Remer's trial.

But within weeks Remer had distributed the report, along with his own annotations, to a wide circle of addresses in Germany. He

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Research was used to "disprove" cyanide use at Auschwitz.

claimed that it had the support of the Max Planck Society. Rudolf halted the distribution by reminding Remer and Herrmann that the report was only for use at the trial.

But keen to publish his report without Remer's political comments — which he claims to find offensive — he began in spring 1992 to seek a publisher himself. Having been turned down by most of Germany's publishing houses, he finally sold the rights to Rüdiger Kammerer, a right-wing extremist, last May.

At that time its publication in Germany was illegal so Kammerer published the report, with a 16-page summary, in Brighton, England. Kammerer claims to have distributed over 10,000 copies.

The Max Planck Society is said to be extremely upset by the affair, and has been working to distance itself from the work since it heard about it last May. It is particularly annoyed, a spokesman says, about claims from right-wing groups that the society supported the report's findings and that Rudolf's dismissal was orchestrated by the Central Council for Jews in Germany — charges which the society vigorously denies.

Last week it issued a statement saying that it supports the German Supreme Court's ruling that the mass murders of Jews in concentration camps is a historical fact that needs no further proof.

A spokesman for the society says that

even if the samples sent to the Fresenius Institute were genuine, Rudolf's interpretation of the data is invalid because there are so many unknown factors involved, such as whether or not the chosen chamber was one of the many known to have been rebuilt before the allied troops entered the camp, or whether residues in the delousing chamber could have remained because much higher concentrations of cyanide were used to kill lice.

The Institute for Solid State Physics dismissed Rudolf without notice on 7 June, saying that he was wrong to use the institute's facilities for work which would bring it into disrepute. It said that Rudolf had misled the Fresenius Institute by implying that the samples were related to work at the Max Planck institute, because he had used institute notepaper.

But Rudolf argues that several people at the institute had known of his work for more than a year before the dismissal, including his supervisor Hans Georg von Schnering, who, he claims, was sometimes supportive. He says the work was done in his free time and the Fresenius institute knew from telephone conversations that his work was independent.

He is bitter about his dismissal, and the publicity around it which has made it impossible for him to find a job. He is also angry that the University of Stuttgart is refusing to consider his submitted thesis, and that no-one will discuss with him the scientific content of his report.

"My only chance now is to prove that I am right", he says, and he is now working with around 20 people from different countries on doing just that.

He refuses to talk about the political implications of his work, claiming that he is only interested in "the truth". Von Schnering admits that he knew of Rudolf's work, but says that the clause in Germany's constitution, guaranteeing freedom of research, prevented him from reporting to senior staff at the institute what Rudolf was doing in his spare time. "Rudolf was a good student scientifically, but the conclusions of the report were all wrong," he now says.

Neither side intend to take the dismissal further. But Rudolf's new allies from the extreme right plan to publish the report in several other languages including English, Japanese and Russian. "Our motivation is to spread the truth," says Kammerer. Last month a German court, in a case brought by the leader of the right-wing National Democratic party, overturned a ruling that the report could not be published in Germany.

Alison Abbott

Factory automation programme ready to fly

Tokyo. The Intelligent Manufacturing System (IMS) project — a global effort to develop the factory of the future — looks set to finally take off this year, five years after it was first proposed by Japan.

The IMS international steering committee, which includes representatives of Japan, the United States, Canada, Europe and Australia, will shortly release a final report recommending launch of a full-scale programme that will run for ten years. Although the report does not include a total cost estimate for IMS, officials close to the programme say it may involve a government-industry investment of as much as US\$4 billion.

Over the past year, a total of 73 companies and 67 universities and research institutes in the five partner regions have participated in 6 test case projects to check the feasibility of a full-scale IMS programme (see *Nature* 362, 97; 1993). Although some problems have been encountered, the international steering committee concludes in the final draft of its report that IMS can be a "win-win opportunity" and a "catalyst" for "unique" international collaboration in industrial research.

IMS was proposed in 1989 by Hiroyuki Yoshikawa, then dean of the faculty of engi-

investment of about \$1 billion in the IMS project. The European Community has more recently estimated that about 4 billion ECU (US \$4 billion) will be required for the proposed ten-year project. But the US Department of Commerce balked at including such a large figure in the final report.

Having looked at the test projects, the steering committee concludes that effective international collaboration was achieved among first-class companies and universities in highly competitive research. "Surprisingly, language and cultural differences did not prove to be a barrier. Rather, these differences added value to the projects and improved their quality in terms of scope, technical content and organization," says the draft.

However, international travel and telecommunications costs proved to be a major burden, constituting 25-30 per cent of project costs. And, even with growing use of computer networks and e-mail, the steering committee expects these overheads to remain high in the full-scale programme, particularly in the initial set-up stage. Intellectual property rights remain a problem — despite extensive guidelines prepared by an IMS committee — and held back some of the projects, as did the different funding mechanisms in each region.

Yuji Furukawa, dean of the faculty of engineering of Tokyo Metropolitan University, who represented Japan on the IMS international technical committee and also seconded for Yoshikawa on the steering committee, expects 3 or 4 of the test case projects to continue into the full-scale programme, and 3 or 4 more may be added. He says they are likely to cover three broad themes: short-term development of global manufacturing techniques; clean "environment-friendly" manufacturing technology; and human and cultural aspects of global manufacturing.

One test case likely to be part of the full-scale programme is "Globeman 21" led by British Aerospace, which aims at determining the most efficient ways of running global manufacturing organizations that have supply chains in different time zones and a range of participating enterprises. Similarly, a project led by Allen-Bradley of the United States with 32 partners in all regions seems set to continue, as does one headed by Mitsubishi Electric of Japan.

However, "clean manufacturing in the process industry" led by ICI in Britain, has ended after failing to win funding from the European Community.

The Japanese have several projects in a domestic IMS programme, which they hope will become part of the full-scale international project. One is a project led by Fuji Electric to develop better computer systems for manufacturing control. According to

Furukawa, whose university is a partner in the project, many US and European companies are very interested in this area which includes concepts such as "agile manufacturing" and "bionic manufacturing". In the latter, the aim is to imitate biological information/production systems, like DNA.

Japanese robot maker Fanuc is leading a project to develop better assembly robots. And Kajima construction company and Hitachi Zosen, a shipping company, are working on "metamorphic" carrying devices that can flexibly adjust to pick up objects of various shapes and sizes.

The full-scale international programme should get underway towards the end of this year, says Furukawa, after the governments of the participant countries and regions have agreed on new terms of reference. The report recommends that the project should be managed by an international steering committee, while regional secretariats and a small inter-regional secretariat will disseminate information, help form consortia, and facilitate review and selection of projects.

David Swinbanks

Fatal blast at experimental reactor

Paris. An explosion killed a technician from the French Atomic Energy Commission (CEA) last week while he was decommissioning an experimental reactor at CEA's research centre at Cadarache in southern France.

The technician, René Allègre, was killed by a blast which caused a 300 m² concrete floor to collapse, also injuring three engineers and another technician, one seriously.

The explosion occurred as they were cleaning residual sodium, slightly contaminated with caesium 137 and tritium, from the cooling circuits of France's first experimental fast-breeder reactor, Rapsodie, which entered service in 1967 and shut down in late 1982.

Early speculation centred on the hypothesis that a violent reaction of sodium with air or water might have caused the explosion — the safety of using sodium as coolant in fast breeders remains controversial. But the first analyses of the accident suggest that hydrogen, produced by the reaction of an alcohol-containing cleaning fluid with the sodium, built-up with unexpected speed.

The French nuclear installation safety authority (DSIN) has rated the accident at two, on a scale of six. Samples taken around the site suggest that no radioactivity escaped beyond the immediate vicinity of the installation, according to the authority.

Declan Butler



Better robots through global collaboration?

neering of Tokyo University, with the backing of the Ministry of International Trade and Industry (MITI). But the proposed central administration by Tokyo received a cool reception in Washington and Brussels, and after several years of discussions the project has evolved along far more decentralised lines, similar in structure to the European computer research programme ESPRIT.

Different funding mechanisms were used in each part of the world for the test projects. In Japan, for example, an IMS centre supported by private companies and MITI provided funds, while the telecommunications and information technology directorate of the European Commission supported some European partners.

Japan's original proposal called for an

"Virtual agency" planned for green research

Washington. The environmental subcommittee of the Clinton administration's new National Science and Technology Council (NSTC) could act as a "virtual agency" coordinating all the federal government's green research plans, a White House-sponsored forum was told last week.

The public forum of administration officials, scientists, environmentalists and industry representatives was the first in a series planned by the Committee on Environment and Natural Resources (CENR), one of nine committees set up by the NSTC to oversee federal research.

Robert Watson, environmental director

for the White House Office of Science and Technology Policy and co-chair of the CENR, says the committee will act as a kind of "virtual agency" to orchestrate federal environmental research.

Peter Raven of the Missouri Botanical Garden in St Louis called the forum "a big success," partly because of the high level of commitment shown by the Clinton administration. Vice President Al Gore, Interior Secretary Bruce Babbitt, Environmental Protection Agency administrator Carol Browner and White House science advisor John Gibbons all spoke at the meeting. Raven says that the NSTC already has established

stronger ties between scientists and policy-makers than did its predecessor, the Federal Coordinating Council for Science, Engineering and Technology (FCCSET).

"This is a more powerful organization than FCCSET," he said, because "the representation is much more comprehensive than it ever was under FCCSET, and it's even more valued by the administration."

But Richard Benedick of the Committee for the National Institute for the Environment, which is pushing for a fully-fledged new agency to fund environmental research, warns that the CENR will face strong resistance from existing agencies — especially if it tries to shift resources from government laboratories to outside scientists.

Watson acknowledges that this is a difficult problem, which he hopes to improve "at the margins" over several years. He believes that a heavy-handed attempt to shift projects from one agency to another would only result in turf battles that would leave "blood all over the floor", and would be a "non-starter" in Congress.

But, he says, there is a real willingness in the White House and in the environmental agencies to make the CENR work. "If I heard a concern [at the meeting], it was that, while we're promoting multidisciplinary research, we may not have the mechanisms to deal with it," he said. "I heard that message loud and clear. We've now got to work for it."

Tony Reichhardt

Pacific sound experiment faces heat

San Francisco. A global study to test the feasibility of measuring climatic changes in ocean temperature using sound has become mired in what one supporter describes as "trial by talk show".

The National Marine Fisheries Service (NMFS), flooded by a deluge of protests from environmentalists and facing likely pressure from some politicians, has delayed implementation of the \$2.9 million Acoustic Thermometry of Ocean Climate (ATOC) project, pending a series of hearings around Monterey Bay beginning April 18.

The experiment, originating at the Scripps Institution of Oceanography, would involve two underwater transducers almost a kilometre underwater off the coasts of Point Sur, California, and Kauai, Hawaii. Once every four hours, the transducers would emit a 20-minute 60–90 Hz coded hum at up to 195 decibels. Eighteen antennae as far away as New Zealand would listen in.

Since sound travels faster in warmer water, the antennas could detect changes in temperature, providing potentially useful information to test theories of global warming.

The dispute erupted last month on MARMAM, the Internet base for marine mammalogists, some of whom feared that the sound would harm or harass whales and dolphins. Two experts on sperm whales — Hal Whitehead of Dalhousie University, Nova Scotia, and his wife Lindy Weilgart — even claimed the noise could potentially wipe out that species. A subsequent story in the *Los Angeles Times* triggered considerable public alarm.

Sperm whales are the only species capable of hearing frequencies below 100 Hz and diving deep enough to be physically injured, according to Christopher Clark, a bioacoustician at Cornell University. But research by Darlene Ketten at Harvard suggests they would virtually have to stick their head into the speakers to be even

temporarily deafened.

Surprised and infuriated by the charges, Scripps officials denied any animals would be killed and said they would halt the experiment at the first sign any mammal was in danger.

Congressman Sam Farr (Democrat, California) says that while the experiment appears to be in the best interest of science, NMFS must address public fears. However scientists involved with ATOC fear the hearings could kill a valuable global warming experiment for no good scientific reason.

Joel Shurkin

French scientists get idiots' guide to PR

Paris. Efforts to achieve a better understanding between scientists and the media have taken a sizeable step backwards with the circulation by the French Centre National de la Recherche Scientifique (CNRS) — Europe's largest fundamental research organization — of an internal guide on how its researchers should handle journalists.

The four-page "CNRS Communications Charter" simply aims to provide researchers with the basics of communication skills, says Nicolas de Schonen, who became head of CNRS's newly-created department of corporate communications last year, after having held a similar post at EuroDisney.

The charter reminds CNRS researchers that as civil servants they are in principle forbidden from publicly criticizing CNRS or government policy. Additionally, it recommends that researchers refer all questions of policy to CNRS management or other better-informed CNRS staff. Researchers themselves, it argues, usually lack a "global view" of a given situation.

De Schonen says his message is that

"it's like within a company, you can't say anything you like". The move raises the question, however, of whether the benefits of avoiding the occasional error or embarrassment outweigh the risks of encouraging the growing trend in many organizations to massage and distort information.

The charter also seems to perpetuate an "us and them" approach to media relations. Researchers are warned, for example, that journalists "need a rapid and definitive response" and may "excessively simplify" their research. Moreover, it advises that researchers "obsess" themselves with "the two or three ideas" they have prepared, and not "preoccupy" themselves "too much" with the questions they are asked. Scientists should "be positive" and "announce good news", it adds.

De Schonen defends this advice as a "provocative" means to make researchers realize that communication is "a two-way process". Corporate communications, adds the charter, can inform scientists of a given medium's "tendency", "interests", and "attitude to CNRS and the topic of research".

Declan Butler

Roche loses round in interferon patent battle

Tokyo. A small Japanese biotech company has won round one of a patent battle with Swiss drug company Hoffmann-La Roche over the former's mass-production of the cancer drug interferon using an unusual technique that employs tens of thousands of hamsters.

After three years of deliberations, the Tokyo District Court has rejected Roche's claim that Hayashibara Biochemical Laboratories, a company of 1500 employees in Okayama west of Osaka — as well as two Japanese drug companies marketing its product — are infringing its worldwide patent for interferon. But Roche is set to appeal against the judgement.

At stake is a sizeable slice of Japan's ¥150 billion (US \$1.5 billion) market for interferon, which is used in the treatment of cancer and viral hepatitis.

With Hayashibara's technique, human lymphoblastoid cells are implanted into newborn hamsters which have been immunosuppressed so as not to reject the cells. The cells proliferate into a large lump which is removed after 3 or 4 weeks, killing the baby hamster. After separation, the cells are suspended in a sterilized culture tank and stimulated by Sendai virus to produce interferon, which is then isolated and purified.

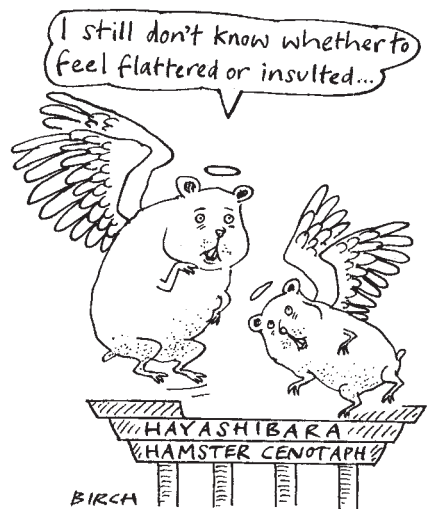
Each year the company sacrifices around 50,000 hamsters in this way, to supply about 10% of Japan's interferon needs. But that share could double if, as expected, the company soon wins government approval to use its interferon in the treatment of patients infected with hepatitis C virus, which is

Praying for baby hamster's souls

Employees at Hayashibara Biochemical Laboratories have an unusual way of purging feelings of guilt about sacrificing tens of thousands of baby hamsters to produce interferon.

Every month, workers from the company line up in front of a memorial cenotaph and a priest chants sutras for the souls of the departed hamsters. While clearly uncomfortable with the relentless killing, company officials point out that with an output of 20 to 40 million international units of interferon per hamster, the sacrifice of each hamster yields enough interferon to treat several human patients.

D. S.



prevalent in Japan and is powering rapid growth of the interferon market (see *Nature* 362, 6; 1993).

Roche's lawsuit claims the Japanese companies are infringing a patent the company filed in 1978 for interferon based on the company's success in separating very pure samples of the material. And Roche claims this is a substance patent that covers Hayashibara's interferon regardless of the process of production or cell line used.

But Hayashibara says that both its production technology and its interferon are its original technology for which it holds patents. The company also rejects Roche's

claim for a substance patent on the grounds that interferon was produced, purified and used for medical treatment long before Roche filed a patent.

Hayashibara portrays the court decision as a victory for a small company against a big multinational that is trying to take control of a basic element of biotechnology. Eckart Gwinner of Roche's Swiss headquarters says it is "obvious" the company will appeal the court's decision. But, given the slow pace of the Japanese court system, the case seems likely to drag on for several more years.

David Swinbanks

Spy interests wrecking Internet security, Congress told

Washington. US government fears about national security are blocking efforts to introduce up-to-date security measures for ordinary users of the Internet and other computer networks, a congressional hearing has been told.

Security problems on the Internet are spiralling, the hearing was told, with the arrival of so-called 'sniffer' programs which go out on the network and search for information, including passwords, giving computer hackers unlimited access to systems linked by Internet.

But Stephen Crocker of the Maryland-based Trusted Information Systems, a computer security specialist, told a hearing on Internet security before the House of Representatives science subcommittee that the US government's wish to stop the proliferation of encryption technology means that Internet users cannot protect their data.

"The government has a strong desire to be able to read the mail," said Crocker. "This takes precedence over security on the

information superhighway." Crocker claimed that the National Institute of Standards and Technology (NIST), a government agency charged with releasing standard security protocols, faced a conflict of interest because it was also supposed to protect encryption technology for government use. "The government priorities are unfortunate, and unless they are changed they will damage computer security and retard the industry," he said.

Industry and government witnesses at the hearing agreed with Crocker that the ability of 'sniffer' programs to sniff out fixed passwords — driving a train through the sole security measure most computer users employ — meant that more sophisticated measures are urgently needed. The sniffers "signal the end of the usefulness of reusable passwords," says Dain Gary of Carnegie Mellon University.

These passwords could be replaced by devices that recognize signatures or fingerprints — or else by far more straightforward

software-based encryption techniques which the government wants to keep for itself. NIST stands accused of wasting time by concentrating exclusively on the former set of techniques. Many in the computer industry believe that signature recognition is a red herring because it requires new hardware which the millions of users already hooked to the Internet do not have and will not buy.

Some witnesses at the hearing were also worried about the likely effects of the government's plans to keep tabs on phone and computer messages by backing a standard encryption chip known as Clipper.

Congressman Rick Boucher (Democrat, Virginia), chairman of the subcommittee, warned that security hazards on the Internet could mean that scientists "cannot be sure that their data will not be tampered with". But witnesses at the hearing acknowledged that most Internet users were oblivious to security risks. This view may change as the number of security incidents on the Internet continues to grow.

Colin Macilwain

Germans face yet tighter rules on lab animals

Munich. German biologists, already working under one of Europe's toughest animal protection laws, may find themselves involved in even more form-filling and delays to experiments if next week's parliamentary agricultural committee approves plans to tighten the law further.

Two clauses are causing particular concern among scientists. One is a requirement for advanced notice of experiments on invertebrates. In other European countries, and in the European Commission (EC) directive on use of animals in research, invertebrates are exempt, but in Germany their status has been unclear because the current law refers to animals, rather than vertebrates, when referring to controls on experiments. Most of Germany's 16 *Länder* interpret the term to mean vertebrates, in line with the EC directive. But some interpret it more literally and require vertebrates and invertebrates to be registered.

The confusion may be cleared up by the proposed changes — but the resolution will not be to the advantage of scientists. While current law requires experiments on animals to be notified two weeks in advance, suggested changes include maintaining the two-week period of advance notice for invertebrates, thus making registration of invertebrate experiments a strict legal requirement, and extending to four weeks the period of notice for experiments on vertebrates.

The second concern is over the proposal that permission should be sought every time an animal is killed for an organ but is not itself used in an experiment. This will not reduce the number of animals killed, science lobby groups argue, but it will hinder research.

Problems began two years ago when Germany, like all other European Union

countries, was obliged to modify animal protection laws in line with EC guidelines. The required changes were benign, but the strong German animal lobby exploited the reconsideration to press for other changes.

Last summer, the state of Baden-Württemberg proposed changes to the

ny's major grant-giving body) have all expressed alarm over the proposed changes, which, they claim, will increase the bureaucracy for scientists without changing the level of protection afforded to animals.

Klaus Fleischmann of the AGF says that although each individual change is apparently trivial, the sum of the changes means that scientists will find themselves burdened with considerably more bureaucracy at a time when the research minister, Paul Krüger, is campaigning for the easing of laws restrictive to science. The parliamentary science committee has also fought against the more restrictive changes to the animal bill.

Fleischmann is worried that the agricultural committee, which has ultimate responsibility for the final draft of the new bill, will not support the science lobby. He says the focus on animal experiments suits the farming lobby by distracting attention from the fact that there will be no requirement to improve conditions for the care and transport of animals intended for food, which are not stringently controlled.

If the wording of the bill is agreed next Wednesday, as planned, it could become law by summer. However, the level of controversy could delay its passage beyond October's general election. **Alison Abbott**

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Animal research: more bureaucracy?

Bundestag sympathetic to the demands of animal rights groups. This was followed by a formal proposal by the Bundesrat to tighten up the animal laws. These proposals have been discussed in several parliamentary committees whose deliberations will be distilled into a formal bill for the Bundestag next week by the agricultural committee.

Research organizations including the Max Planck Society, the AGF (which represents national research centres) and the Deutsche Forschungsgemeinschaft (Germa-

Soros award for 24

Moscow. Twenty-four Russian scientists have been certified as George Soros Emeritus Professors as part of the International Soros Science Education Program (ISSEP) to support education in physics, mathematics, chemistry and biology in the former Soviet Union.

Besides providing the recipients with stipends of \$200 a month for five years, the titles, awarded personally by Soros on a recent visit to Moscow, seem likely to become a valued mark of honour.

Among the recipients were Evgeny Feinberg, a physicist and a close friend of the late Andrei Sakharov, and Daniil Lyebbedev, the well-known fighter against Lysenko during the darkest of times. One laureate did not live to see the awards ceremony, the former president of the Russian Academy of Sciences and director of the Kurchatov Institute, Anatoly Aleksandrov.

The ISSEP is intended to fill the gap left by the two existing Soros programmes supporting basic science and the humanities. The programme is planned to last for five years. Soros has already allocated \$25 million for the first year.

Besides these elite awards, grants are being established for high-school teachers and university lectureships. Three thousand science teachers will receive grants; five hundred university professors will receive the title of Soros Professor. **Vladimir Pokrovsky**

Czech institute to start from scratch

Prague. Plans to create a learned society in the Czech republic may go ahead, despite the failure of the government to sanction the idea with a law.

Rudolf Zahradník, the president of the Czech Academy of Sciences, wants to create a new learned society in line with most Western European countries. Last spring he asked the Czech president Václav Havel to elect the first 20 members of the proposed society. Havel agreed, on condition that the founding of a society was supported by a law.

But the government has steadfastly refused to create such a new law, resisting what it sees as overlegislation in reaction to the country's communist past. A recent interministerial meeting with scientific advisers once again gave the idea short shrift.

Zahradník is disappointed that the chance for a launch of the proposed society by Havel, which would have lent it a strong

degree of credibility and idealism, has been lost. But he still hopes that the learned society will be formed this year from a core of academics who have already joined a Foundation for the Reestablishment of a Learned Society.

In communist times, the Czech Academy acted as a learned society as well as controlling around 80 basic research institutes. But its role as a learned society was corrupted by politics, and the Czech Academy was the only academy in all of the former communist states in eastern Europe to abandon this function, after the 'velvet revolution' in 1990. Now it only runs the research institutes.

Zahradník would like the new learned society to take over the title Czech Academy of Sciences. This would minimize confusion, being in line with learned societies in other countries. **Alison Abbott**

French research moves to meet public needs

Paris. The French biomedical research organization, INSERM, has launched a major initiative to speed up the research community's response to society's requirements.

The initiative is designed to meet growing public demand that science should meet the needs of society. The result will be a new service of "collective expertise" that will organize task forces of researchers to produce quick answers to contemporary questions of public health, as well as analyses of potential health hazards and opportunities.

Philippe Lazar, director-general of INSERM, says that the new service is aimed at "short-circuiting" the traditional research system.

The research timescale, he argues, creates a delay in the integration of current knowledge into the thinking of governments and companies. "It's now often more urgent

to take stock [of existing knowledge] than to launch new research," he says.

Lazar's idea is that researchers are best placed to sift through the avalanche of increasingly complex scientific papers in search of trends that are pertinent to decision-makers. In practice, a new office within INSERM's department of partnership for economic and social development will coordinate a series of *ad hoc* multidisciplinary task forces composed of INSERM researchers and outside experts, which will produce reviews on political, economic, and industrial issues within six months of a request.

Researchers who chose to participate in the initiative will be paid for doing so, at teaching rates, on top of their normal salaries.

INSERM has already carried out studies, for example, on the effects of low-frequency radiation on health (commissioned by

Electricité de France), the health hazards for pregnant women of VDUs, and which contraceptive pills should be available on prescription. One of the first contracts to be carried out by the new service will be an assessment of the research strategy of a French pharmaceutical company.

Customers will be charged between FFr 300,000 and FFr 500,000 for such services. But Lazar emphasizes that "collective expertise" is not a profit-making consultancy. The charges will simply cover costs.

One possible conflict, admits Lazar, is between INSERM's duty to publish, and the demand of some customers for confidentiality. He says that INSERM will negotiate for as much open access as possible, but that in some circumstances, companies, for example, may be allowed exclusive access to an expert opinion for a certain period before it is made public.

The new service will also produce expert opinions on its own initiative. Part of the thinking behind this is to provide an early warning system. Indeed, INSERM officials say it is partly a response to the acknowledged failure of the French research community to recognize and respond to the danger posed by the contamination of French blood supplies in the early 1980s.

The first such study will be an assessment of potential health risks of human products used in therapeutics. Lazar reckons that such prospective studies will not only anticipate issues of public health but will also provide fresh insights into INSERM's own research strategy.

Declan Butler

Competitiveness threat "underrated"

Brussels. The "sheer scale" of the threat of loss in Europe's industrial competitiveness has been "largely underestimated", says a group of leading European industrialists. The message comes in a report to the European Commission by the Industrial Research and Development Committee of the European Commission (IRDAC), an advisory committee made up of leading industrialists. It called on the European Union (EU) to encourage greater collaboration between industry and educational institutions to counter the threat.

The report, *Quality and Relevance, the Challenge to European Education*, will add momentum to Jacques Delors' call in his recent white paper on 'competitiveness, growth and unemployment' (see *Nature* 386, 599; 1993), for Europe to put greater emphasis on education and training.

The council of ministers and the European Parliament will soon consider approving a massive increase in funding for two new EU education programmes. The commission has proposed increasing spending by two-thirds to ECU1.01 billion (US\$1.16 billion), within a new programme called SOCRATES. It also wants to double support for vocational training schemes to ECU801.8 million.

Yves Farge, the chairman of IRDAC, says that Europe's educational institutions are not adapting quickly enough to "the third industrial revolution". To counter this, IRDAC urges the EU to increase collaboration between industry and education in its programmes.

In particular, it recommends creating more apprenticeships, generalizing student and staff placements in companies throughout the EU and the sharing of training materials. A commission official

says that the EU intends to fund consortia made up of companies and educational institutions to develop such projects.

Among other things IRDAC urges the EU to do is to encourage mutual recognition of educational qualifications among members and arrange that students living in one member state will be entitled to comparable education in any other. IRDAC also wants the EU to promote foreign language learning and distance education, and collect statistics on education and training.

Declan Butler

Spanish science plan stresses applications

Barcelona. A new strategy for publicly funded science in Spain to come into effect at the end of this year is expected to emphasize applied and multidisciplinary research.

The new blueprint — which will give a much bigger role to the Ministry of Industry — is being prepared by the secretariat of the National Plan for Research and Development (NPR&D) at Spain's Ministry of Education and Science and has so far not been discussed outside the government.

Explaining the proposal, the research secretary, Elias Fereres, says: "We have to make the transition from investigation in small specialized groups to wider interdisciplinary groups, more integrated with the public sector." Fereres denies that the new plan will kill off basic research. "It is only a question of placing the incentives elsewhere," he says. But he accepts that some established lines of research "are bound to suffer".

Another controversial aspect of the plan is its exclusion of specific programmes for the autonomous regions. One region — Catalonia — is pushing for power-sharing schemes which the central government is not keen to

concede. The new plan would enhance an advisory council of the autonomous regions, which up to now has been inoperative.

Fereres says the plan will give support to public-sector industry, which is the primary investor in research and development in Spain. Favoured sectors include agro-business, food, environment, materials science, transport, reforestation, and industrial electronics.

The new plan tries to counter criticism that the current NPR&D has failed to set clear priorities for the application of research to industry. One proposal for dealing with this was the creation of a new agency for technology transfer, but this was discarded due to the difficulties in trading-off the interests of the ministries concerned.

Current thinking favours generating demand from industry. A new agency will help the NPR&D to detect areas of industrial interest where multidisciplinary groups of scientists and researchers could act. The first version of the plan will be ready before summer. By autumn, the research programmes will be open for bids.

Luis Angel Fernández Hermana

Chalk River got there first

SIR — Carlo Rubbia's proposal for an accelerator-driven energy amplifier¹ seems to have overlooked earlier work here at the Chalk River Laboratories of Atomic Energy of Canada Ltd (AECL). What Rubbia and, separately, Los Alamos propose is a version of a concept studied here for more than four decades, but which has hitherto been rejected on economic grounds.

The original concept was due to W. B. Lewis² in 1952, and entailed the use of spallation neutrons, generated by the bombardment of heavy metals by high-energy protons, to convert fertile into fissile fuel. Spallation neutrons had then only recently been reported³⁻⁴, but the first experiments at Chalk River used cosmic-ray protons to measure spallation yields. Lewis⁵, in the early 1960s, then advocated the Intense Neutron Generator (ING), based on a 1-GeV, 65-mA proton beam from a linear accelerator and a molten lead-bismuth target, as a source of neutrons for many research applications⁶.

Work on fuel breeding continued in parallel. Lewis⁵⁻⁶ discussed the economics of thorium conversion into fissile ²³³U using the ING design as a source of neutrons. But funding for ING was not forthcoming; the project was cancelled in 1968.

Interestingly, it was appreciated from the start that reprocessing could be an issue. A report⁷ in 1973 noted that Canadian CANDU reactors would run as well on ²³³U from thorium as on ²³⁵U from natural uranium and suggested that thorium fuel pins could be directly irradiated by spallation neutrons. The same document also raised the possibility that a "sophisticated fuel management scheme" might allow the return of intact but partially spent fuel pins to the reactor after upgrading by the spallation source.

It was also recognized that the energy cost of accelerating protons could be at least partly offset by the heat generated in the spallation process. It was estimated⁷ in 1973 that 62 per cent of the energy required to operate the accelerator could be derived from heat recovered from the target. In the early 1980s, consideration of alternative target designs suggested⁸ that the spallation heat-gain could be as much as a factor of five, which is great enough to allow all the accelerator power to be obtained from the target heat. In the course of the later ZEBRA project⁹, a conceptual design optimized for fuel conversion suggested that there would nevertheless be a surplus of 110 MW of electrical power from the target assembly.

Apart from the work at Chalk River, the notion of electronuclear breeding has been studied at Livermore¹⁰ (in the 1950s) Brookhaven¹¹ and at Los Alamos¹² in the

United States, as well as in Russia and Japan.

M. Harvey

J. Unglin

*Atomic Energy of Canada Limited,
Chalk River Laboratories,
Chalk River, Ontario, Canada K0J 1J0*

1. *Nature* **366**, 392 (1995).
2. Lewis, W. B. *The Significance of the Yield of Neutrons from Heavy Nuclei Excited to High Energies*, AECL 968 (1952).
3. Goeckermann, R. H. & Perlman, I. *Phys. Rev.* **73**, 1127 (1948).
4. O'Connor, P. R. & Seaborg, G. T. *Phys. Rev.* **74**, 1189 (1948).
5. Lewis, W. B. *How Much of the Rocks and the Oceans for Power? Exploiting the Uranium-Thorium Fission Cycle*, AECL 1916 (1964).
6. Lewis, W. B. *The Intense Neutron Generator and Future Factory Type Ion Generators*, AECL-3190 (1968).
7. Fraser, J. S., Hoffmann, C. R. J. & Tunnicliffe, P. R. *The Role of Electrically Produced Neutrons in Nuclear Power Generation*, AECL-4658 (1973).
8. Fraser, J. S., Hoffmann, C. R., Schriber, S. O., Garvey P. M. & Townes, B. M. *A Review of Prospects for an Accelerator Breeder*, AECL-7260 (1981).
9. Schriber, S. O. *The ZEBRA Program at CRNL*, Proc. 1981 Linear Accelerator Conference, Sante Fé, New Mexico, October 1981. Published as LA-9234C by Los Alamos National Laboratory (1981).
10. Armagnac, A. P. *Popular Science*, 108 (1958).
11. Grand, P., Batchelor, K., Powell, J. R. & Steinberg, H. *The Accelerator Breeder, An Application of High-Energy Accelerators to Solving Our Energy Problems*, IEEE Trans. Nucl. Sci. NS-24, 1043 (1977).
12. Lawrence, G. P. *New Applications for High-Power Proton Linacs*, Proc. 1990 Linear Accelerator Conference, Albuquerque, New Mexico, September 1990. Published as LA-12004C, by Los Alamos Laboratories, 1991.

French blood test

SIR — After reading your recent articles on the "second" French blood affair¹, based on articles I wrote for *Libération*, I would like to make two corrections.

You write: "While Seytre claims that the French test was unreliable until June 1985, Françoise Brun-Vézinet of the Claude-Bernard hospital in Paris — whose data are cited by Seytre — says that, although there was some variability between batches, the quality of the test itself was not a problem. . . . 'We were still testing it in April', says Brun-Vézinet."

What I actually wrote was: "Until April 1985, the Diagnostics Pasteur test gave a high number of false negative results." I also specified that the defective batch was batch no. 86 and added, "all the kits produced by Diagnostics Pasteur at that time were not as poor . . . the quality varied significantly from one batch to another." Rather than contradicting what I wrote, the quotation from Brun-Vézinet confirms it.

I am also concerned that you identify me as "the translator of Robert Gallo's book". This translation is only one of seven translations I have done and it is completely irrelevant to my work as a scientific journalist and to my recent articles on the blood scandal.

The articles I wrote and to which you refer are direct products of research that I have been conducting over the past several years on the history of AIDS in association with INSERM Unit 158 (the French National Institute for Health and Medical Research), currently with a grant from the ANRS (National Agency for AIDS Research). Moreover, these articles are a continuation of the work I did for a book published last May on the history of AIDS research².

Bernard Seytre

*INSERM 158, 149, rue de Sèvres,
75015 Paris, France*

1. *Nature* **367**, 673 (1994).
2. Seytre, B., *Sida: Les secrets d'une polémique. Recherche, intérêts financiers et médias*, prefaced by Pr. Willy Rozenbaum (Presses Universitaires de France, Paris, 1993).

Not so puzzling

SIR — Jorge L. Sarmiento¹ notes that there has been an unprecedented slowdown in the accumulation of atmospheric CO₂ and searches for geophysical answers to the puzzle because "fossil fuel emission, estimated from the most recent United Nations Statistics" show a slight increase and " . . . there is no reason to believe that the past year and a half were unusual".

The past few years have indeed been unusual. The former second biggest fossil CO₂ emitter — the Soviet Union — has undergone drastic contraction, along with the loss of additional large CO₂ contributions from the Central/East European countries. Sarmiento really has unwarranted faith in the data, which from these areas during this time have been of questionable reliability. Alternative estimates from recent BP data², using coefficients previously developed with the aid of industry sample analysis³, indicate that the decline was dramatic enough to reverse by 1991 the previous rapidly rising trend in global emissions. Global (fossil-fuel) CO₂ emissions in 1992 were still below the peak 1990 levels (the estimates for global total emissions being 6,037, 6,081, 6,014, and 6,028 million tonnes of carbon for the years 1989–92 respectively).

Deforestation is also a significant source and although data are notoriously unreliable there are indications of a decline driven by dramatic reductions in Brazil since 1988. There may be no need to hunt for geophysical causes; let's try and check the emissions data first.

Michael Grubb

*Royal Institute of International Affairs,
Chatham House, 10 St James's Square,
London SW1Y 4LE, UK*

1. *Nature* **365**, 697 (1993).
2. *BP Statistical Review of World Energy* (BP, London, 1993).
3. Grubb M. et al. *Energy Policies and the Greenhouse Effect, Volume II: Country Studies and Technical Options* Appendix 2. (Dartmouth/RIIA, Aldershot, 1991).

Trust the wildlife volunteers

SIR — The amendment to prohibit the US National Biological Survey (NBS) from accepting the services of volunteers (*Nature* 367, 400; 1994) was supported by two arguments in the House of Representatives (*Washington Post*, 25 December 1993). They were that volunteers are incompetent and that they are biased. That is not the European experience.

I write as director of the British Trust for Ornithology and chairman of the European Bird Census Council. In Britain and Ireland, the volunteer network has made invaluable contributions to the assessment of wildlife resources. This is especially true in ornithology, where we have had two distributional atlases of breeding birds at a 20-year interval (with an atlas of winter distributions intercalated), annual population monitoring of most common and many rare birds, and continuing assessment of breeding success and survival. From this information, through an integrated programme of analysis and through related studies on how birds use their habitats and how they are affected by management practices, we can identify species, guilds and habitats that seem to be in trouble and the possible causes of the problems. This in turn can lead to more specific and intensive studies, by volunteers or professionals, to illuminate the causes further and to provide the basis for sound conservation action.

This volunteer-based work is accepted as a reliable foundation for policy-making and planning by government, by conservation bodies and by the agricultural, forestry and other industries. The competence of the amateur birdwatchers is recognized to be as high as that of professionals. Bias is not an issue, not only because there is no empirical evidence for such bias but also because the participants recognize that biased evidence would not allow conservationists themselves to develop sound policies. (In addition, it may be significant that most of the work is not organized by bodies that are also involved in conservation campaigning.)

Although Britain may have, perhaps through no more than historical accident, the best developed and most highly integrated volunteer-based schemes for ornithological monitoring, there are similar programmes in other European countries. Indeed, the European Bird Census Council is promoting the use of such schemes as ways of monitoring the health of the wildlife environment throughout Europe.

Programmes to use volunteers for wildlife assessment are not as developed or as integrated in North America as in Britain and Ireland. The establishment of the NBS represents an opportunity for a great

leap forward. If the opportunity is to be seized, the key message to be taken on board, by politicians and by professional scientists alike, is "Trust the volunteers".

Jeremy J. D. Greenwood

*The Nunnery,
Thetford, Norfolk IP24 2PU, UK*

Fraunhofer funds

SIR — We appreciate your interest in the Fraunhofer Institutes. Unfortunately, a passage in Alison Abbott's article (*Nature* 365, 481; 1993) is open to misinterpretation. She reports that the German Federal Research Minister, Dr Paul Krüger, had said to me privately — in contradiction of his public statements — that he would like to cut the funds dedicated to fundamental research from 40 per cent to one-third, in favour of the Fraunhofer-Gesellschaft. On the contrary, the minister stressed the positive value of application-oriented research, without mentioning any redistribution of funds. And I myself pointed out in my conversation with Abbott that it would be desirable for more to be done for applied research, which is the very basis of innovation, and thus serves to enhance the competitiveness of German enterprise within international markets. In view of the meagre funds available, the question of their appropriate distribution is paramount.

Hans-Jürgen Warnecke

(President)
*Fraunhofer Society,
Leonrodstrasse 54,
D-80636 München, Germany*

Smallpox bargain

SIR — Your leading article "End of the line for smallpox virus?" (*Nature* 366, 711; 1993) was responsible and balanced but I wish to correct the statement that the eradication of the disease was costly. According to a detailed study to ascertain all the international costs, including estimates to account for the services of the Peace Corps and Oxfam volunteers, and the contributions by several organizations of vaccines and vehicles, the total was just under \$100 million. In contrast, the cost to the United States in 1969 for smallpox vaccination, complications of vaccination, quarantine and so on amounted to \$150 million — for one year alone.

These comparisons show that smallpox was eradicated, by an international effort, for a cost equivalent to two-thirds of the expenditure by the United States in eight months for smallpox control. This seems to me to be a bargain. The efforts by the World Health Organization and those

organizing the programme, Drs Henderson, Fenner and Arita, deserve more than the somewhat snide remarks about the contributions of this often-criticized body. To eradicate a major disease deserves unreserved commendation and was indeed recognized by the award of the Japan Prize to several of the principals involved in the programme.

Fred Brown

*US Department of Agriculture,
Plum Island Animal Disease Center,
PO Box 848,
Greenport, New York 11944-0848, USA*

Avoiding bends

SIR — Daedalus has expressed the idea (*Nature* 365, 18 & 112; 1993) that seals and dolphins must have evolved an inhibitor against air bubbles, leading to the considerable resources of DREADCO's research team being channelled into the search for the responsible protein. However, the members of the research team of their corporate adversaries, DREAMCO, were aware that such a search was in vain. They know that seals and dolphins avoid the bends by denying nitrogen the opportunity to become supersaturated in the blood. Unlike man, dolphins and seals do not breathe while submerged and forcibly exhale before a deep descent. Hence there is no nitrogen available to supersaturate the blood.

Seals and dolphins are able to stay submerged for great lengths of time because of large blood volumes and superb oxygen-carrying capacity. The researchers at DREAMCO are now working to impart this oxygen-carrying capacity to humans. The genetic code responsible for the production of haemoglobin is being spliced into bacterial cells. The 'super oxygen' bacteria are then ingested in a similar manner to an asthmatic spray. With the help of the bacteria, the code makes its way into blood cells and begins to produce vast amounts of oxygen-carrying haemoglobin. These infected cells are unable to reproduce and eventually die, ensuring a continuing market for DREAMCO's Aerobic Spray.

The market for such a product would appear to be enormous. The benefit for sporting people is clear. Other benefits would include a decrease in the number of drownings, due to the ability of a user to forgo fresh oxygen for up to five times as long. Indeed, some believe DREAMCO may already be successfully trying out the spray with the company-owned professional cycling team which has won every event entered over recent weeks.

Vincent Craig

*Department of Chemistry,
Australian National University,
Canberra, ACT 0200,
Australia*

Wildlife conservation as wealth

Valerius Geist

Wildlife conservation is incompatible with global markets or private ownership. What is needed is a 'tribal' system of management such as that in North America that creates both wealth and jobs while sustaining resources.

How can wildlife be conserved while creating wealth and employment? This recurrent topic of debate once again surfaced in last month's meeting of the Convention on International Trade in Endangered Species (CITES) in Geneva to coordinate programmes for saving the world's tigers from extinction. Indeed, the question has recently been discussed by the International Union for Conservation of Nature (IUCN) and its specialist advisory group, at the First International Wildlife Management Congress in September 1993 in San Jose, Costa Rica, by the World Bank and by the authors of several academic and popular books on wildlife and biodiversity¹⁻⁴. Yet a bad initiative continues to be pursued — the creation of a global luxury market in wildlife products. As we shall see, this approach to sustainability is invariably ineffective, with devastating consequences for conservation and detrimental effects to agriculture and public health.

Why is there never any mention of an alternative approach — the 70-year-old North American system of wildlife management? Although deliberately kept outside a global consumer market, this system generates both wealth and employment and is a proven conservation success — a model example of sustainable ecological development. It has returned wildlife from the brink of extinction, made it abundant and economically important⁵⁻⁸ and, in the process, defeated Garrett Hardin's "Tragedy of the Commons"⁹ — the inevitable depletion of public resources when exploited freely without effective societal control. (The concept was throughly misinterpreted in the *Economist*¹⁰ in an unsigned article on public ownership of wildlife in North America.) The system illustrates how a renewable resource, publicly owned and managed, can be exploited by the private sector to create a job-sensitive manufacturing and service industry worth more than \$70 billion annually in the United States and Canada. (This compares to a global market in wildlife products, legal and illegal, of only \$8 billion¹¹⁻¹³.) Restraints, sternly and intelligently applied through political consent to Adam Smith's "unseen hand", can therefore generate wealth and jobs while sustaining wildlife resources. Here is a

rare case of 'doing well by doing good', and it is one of North America's great cultural achievements.

The important lesson is to keep wildlife out of the marketplace, and thus out of private hands, while encouraging its diverse use under close public scrutiny. Like the US automobile industry, which generates wealth and employment from an inefficient but convenient form of transportation, so North American wildlife conservation generates wealth and employment by deliberately rendering the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Symbol of success — the North American wapiti.

public use of wildlife as inefficient as possible. This inefficiency is legislated: state, provincial and federal restraints on how wildlife may be killed for consumption, or even observed, are directly responsible for the many varied market opportunities that the private sector can exploit. The result is a rich, innovative, job-intensive manufacturing and service industry that supports the tightly controlled use of public wildlife. Unlike conventional economics, which brings the resource to the consumer, North American wildlife economics takes the consumer to the resource.

The traditional core policies of North American wildlife conservation remain largely intact: the public ownership of wildlife with bans on the private ownership of native species; a ban on trade in dead wildlife (excluding furs); an annual egalitarian allocation of surplus wildlife by legislatures after consultation with the public; a ban on the killing of wildlife without good reason; a ban on the waste of

wildlife after killing; and the involvement of scientists and wildlife professionals in management.

To show how the system works in practice, consider what you would have to do if you wanted to eat grouse. Under the legislation, you could not buy it from a butcher's shop or hire someone else to kill it for you. Rather, you must go to the expense and trouble of equipping yourself for the hunt, learning the art of hunting and its legal aspects, and paying for transport and clothing as well as for the storage

of any game that you do not want to eat immediately. Such a requirement demands specialized goods and services. Similarly, if you were a bird-fancier, you would not be able to buy a native species to admire in a cage. Instead, you would have to spend money on the binoculars, clothing and transport required to watch birds in their native habitats. The result? A lively industry in support of bird-watching. Legal restrictions on the kinds of hunting and fishing gear that may be used, although introduced to deter people from participating in these sports, can encourage the development of new market niches. For instance, under legislation, deer have been free for the taking but only with bows and arrows.

The result has been an innovative manufacturing industry in archery equipment, private archery ranges, specialized camouflaged clothing and novel lures to attract the quarry, and a deluge of magazines, books and tapes to keep the hobbyist well informed.

A lot of money is spent by people who hunt, fish or view wildlife in North America. Wildlife, available to anyone for a nominal licence fee, is in reality acquired only after great expenditure. Wyoming, for instance, makes about \$1 billion annually from its 960,000 big-game animals. That is, each living big-game animal contributes about \$1,000 a year to the state's economy. Although most of this income comes from viewing wildlife, the viewers and the consumers largely overlap.

North American wildlife conservation is in principle analogous to the management of wildlife by native tribes, such as the Labrador Cree in Canada or some tribes in Africa^{2,14,15}. Citizens may exploit wildlife, provided that they follow certain

Martin W. Grosnick/Ardea

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Watching for California condor at Los Padres National Forest.

legal rules and regulations; and they may express their opinions on wildlife conservation, individually or through conservation organizations, to management agencies or their elected representatives. Wildlife conservation becomes a very public affair; all may participate.

It was not always so. At the turn of the century, North America's wildlife suffered in full from the 'tragedy of the commons'. Uncontrolled markets in wildlife luxury products and a last-ditch attempt by the military to subjugate 'hostile' natives by starving them into submission virtually eliminated wildlife from much of the continent. In the decade before and during the First World War, a system of conservation soon developed that has not only returned 'hunnable' wildlife to abundance but also protects nature through public stewardship. Wildlife was made a source of wealth and employment, despite big-game animals being outnumbered by firearms by at least 8 to 1, by people by 9 to 1 and by livestock by more than 120 to 1. Meanwhile, rich land-owners continue to test the system in court, hoping to gain control over the public's wildlife for private gain^{16,17}.

The 'tribal model' of wildlife conservation that excludes trade and private ownership has worked elsewhere too. Versions are found in Scandinavia, Switzerland, Mongolia, Japan and Africa. The system rapidly became a success in the North Western Territories of Canada when native people were empowered to take over the management of their wildlife, as it did when wildlife was returned to tribal control in several African states¹⁻⁴. In North America, there have been great differences in the system's application and results.

By contrast, a global luxury market in wildlife is compatible with neither conservation nor good economics. Markets in dead wildlife reward its killing, foster its private ownership and manipulation,

spawn disease and are open to corruption¹⁸. With the spread of 'game ranching' in North America, mongrel escapees from deer farms are causing genetic pollution of wild stocks of the giant deer, or wapiti. And game farms have spread diseases to livestock, as exemplified by the spread of bovine tuberculosis in North America (in New Zealand it threatens the country's dairy exports¹⁹). Further, trade in wildlife, legal or otherwise, is endangering not only agriculture but also public health²⁰. Although transmissible spongiform encephalopathy circulates as 'mad cow disease' in captive cattle herds in the United Kingdom, in North America it is found in three species of free-living deer²¹⁻²³, the legal and illegal source of stock for deer farms. Escaped animals may form feral populations and displace native ones, as did the sika deer in North America; predators are unwelcome in most commercial game-raising ventures^{24,25}. Finally, in the face of high market demands, illegal killing and selling of public wildlife is both lucrative and reasonably safe, and an industry in wildlife products remorselessly pits wealthy

private interests, backed by powerful agricultural bureaucracies, against public conservation interests — both in and out of the courts²⁶.

Private interest in tropical forests and ocean fish, and the market price on them, have surely been powerful factors in the demise of these resources. So why place this price on dead wildlife too? The IUCN's approval of global trade in wildlife, and the weakening of CITES control of trade in endangered species, would make the IUCN a sorcerer's apprentice. Uncontrollable forces of wildlife destruction would be unleashed, and the owners of large private ranches would seize public wildlife as a private asset. The oversimplified arguments² for legalizing the ivory trade in South Africa have already gained a credence that is disturbing: the resumption of a limited ivory trade is now being proclaimed by some spokesmen for South African game ranchers as "the solution" to nothing less than "sustainable development", generating income for impoverished local communities. But have no doubt, a 'split listing' for the marketing of raw ivory would be the thin end of the wedge, providing a renewed incentive for poaching and stimulating a dangerous new demand for ivory in growing economies such as China. The sorcerer's apprentice is already on the loose in several Canadian provinces and US states, where the rush towards game ranching has gutted wildlife conservation legislation, despite much adverse publicity and — largely futile — public opposition to powerful private interests. And all this while wildlife conservation in North America evolved towards minimizing the net value of its wildlife resource²⁷. When will politically powerful conservation organizations learn from the experience of others and stop endorsing a global luxury market in wildlife products? □

Valerius Geist is in the Faculty of Environmental Design, University of Calgary, Calgary, Alberta, Canada T2N 1N2.

- Schaller, G. *The Last Panda* (University of Chicago Press, 1993).
- Bonner, B. *At the Hand of Man* (Knopf/Simon & Schuster, 1993).
- Wilson, E. O. *The Diversity of Life* (Harvard University Press/Viking, 1993).
- Proc. 1st Int. Wildl. Management Congr., Costa Rica, 19–25 September 1993 (The Wildlife Society, Bethesda, Maryland, in the press).
- Geist, V. *Conserv. Biol.* **2**, 1 (1988).
- Trans. 54th North Am. Wildl. natn. Res. Conf., 171 (Wildlife Management Institute, Washington DC, 1989).
- Brown, R. D. *The Biology of Deer*, 554–561 (Springer, 1992).
- Commercialization and Wildlife Management: Dancing with the Devil* (ed. Hawley, A. W. L.) (Krieger, 1993).
- Hardin, G. *Science* **162**, 1243 (1968).
- Economist*, 21 (22 October 1988).
- 1991 US Fish and Wildlife Service National Survey of Fishing, Hunting and Wildlife Associated Recreation (US Fish & Wildlife Service, Washington DC).
- 1991 Canadian Wildlife Service Survey (Canadian Wildlife Service, Department of the Environment, Ottawa, Ontario, 1991).
- US TRAFFIC Office (Washington DC, 1991).
- Feit, H. A. *Swedish Wildl. Res. Suppl.* **1**, 24 (1987).
- Dueck, H. A. *Carnivore Conservation and Interagency Cooperation: A Proposal for the Canadian Rockies* (MSc thesis, University of Calgary, 1990).
- Morill, W. I. & Buckmaster, R. *Trans. 58th North Am. Wildl. natn. Res. Conf.* (Wildlife Management Institute, Washington DC, in the press).
- Hargrove, E. C. *Foundations of Environmental Ethics*, 57–63 (Prentice Hall, 1989).
- Poten, C. *Natn. Geographic* **180** (3), 106 (1991).
- Bovine Tuberculosis in Cervidae: Proceedings of a Symposium* (ed. Esset, M. A.) (US Dept of Agriculture, Miscellaneous Publication No. 1506, 1992).
- Fanning, A. & Edwards, S. *The Lancet* **338**, 1, 243; 1, 253 (1991).
- Williams, E. S. & Young, S. J. *Wildl. Diseases* **16**, 89 (1980); **18**, 465 (1982).
- Guiry, D. D., Williams, E. S., Yanagihara, R. & Gajdusek, D. C. *Neurosci. Letts* **126** (2), 195 (1991).
- Guiry, D. D., Williams, E. S., Yanagihara, R. & Gajdusek, D. C. *Acta Neuropathologica* **81** (5), 475 (1991).
- Pruitt, W. O. Jr *Explorer J.* (30 March 1985).
- McNabb, J. *Can. J. Zool.* **69**, 2, 283 (1991).
- Dorrance, V., McCarthy et al. (No. C90-0110; 1991); *ibid* (Civil Actions No. 6244/6245); ruling by Wyoming Game and Fish Commission, 26 February 1993.
- Lueck, G. J. *Legal Stud.* **18** (2), 291 (1989).

Modelling high- T_c superconductivity?

The ubiquitous Ising lattice has been used to model competitive ordering forces such as may account for some of the properties of the copper-oxide planes in ceramic superconductors.

The Ising lattice is not so much a model as a notation, it now appears. Originally, in the 1920s, the idea was to make a model of magnetic materials, ferromagnets for example. The procedure was straightforward: represent the structure of the solid by an appropriate geometrical lattice, imagine that there is a magnetic dipole at each of its vertices, define some rule for calculating the total energy of the system from the orientations of the magnetic dipoles at all possible pairs of vertices and then treat the problem as one in statistical mechanics. The hope was that it would be possible to show that such a model would behave as a real ferromagnet does; in other words, that there would be a critical temperature (the Curie temperature) below which magnetic order would be the rule, above which it would not be.

The frustration of this ambition has been amply described. In general, even with simplifying assumptions about the energy of interactions between differently located lattice sites, real three-dimensional problems are not soluble. Even in two dimensions (as with crystalline solids in a plane, Langmuir-Blodgett films perhaps), only drastic simplification will allow analytical solutions of model problems. One powerful simplification is to suppose that only magnetic elements at literally nearest-neighbour vertices contribute to the interaction energy. (On a square lattice, for example, interactions across the diagonal would be counted as zero.) Then, most order-disorder problems on reasonably shaped lattices can be solved exactly.

But do not the simplifying assumptions empty the baby with the bath water, so to speak? A real ferromagnetic material, say a crystal of iron, cannot really be discussed as if it were nothing but a collection of magnetic dipoles sited on the vertices of a geometrical lattice, each of which interacts energetically only with its nearest neighbours in a strictly classical fashion. Indeed, atoms in such a crystal acquire magnetic moments (other than those of their nuclei) by losing electrons to the Fermi sea, which is the means by which the crystals are able to conduct electricity.

The obvious snag is that the ions left behind continue to interact with the electron sea collectively, which means that, at the very least, the problem of ferromagnetism boils down to that of calculating the expectation value of the total electron spin for a system of atoms capable of being ionized and which are located on the vertices of a regular lattice. That is a problem in quantum mechanics, which leads to what is called the

Hubbard lattice, a system of quantum spins in a regular array. To tell from the literature, people are still enumerating the eigenstates of such a system. It is even less soluble than the Ising lattice.

That does not mean that the Ising lattice has been of no value. On the contrary, it is a literally exact model of order-disorder in, say, α -brass (Zn to Cu in the ratio 3:1). Measurements show that, in the ordered (and hardened) state, extra energy is needed to melt to solid than when the atoms are randomly arranged. The other side of that coin is that it is possible to measure the latent heat of the phase transition for the disordered to the ordered state when a sample in the disordered solid is annealed. That is something tangible, technically a first-order transition.

The more substantial interest of the Ising lattice has been, over the past several decades, heuristic, as the saying goes. Allowing for the simplifications inherent in any attempt at analytical solutions, it nevertheless emerges clearly that three-dimensional systems allow of first-order phase transitions, the simpler two-dimensional systems allow of only second-order transitions (where there is a specific heat anomaly, but no latent heat, at some critical temperature) and that simple one-dimensional systems in which only nearest neighbours interact do not allow order-disorder phase transitions in either sense. (They cannot, for a single break-point in a linear lattice will destroy long-range order.) None of that is surprising, but it is good that it is confirmed.

By now, there seems to be no limit to the uses of this simple model. It is, for example, a natural description of adatoms on a solid surface (but with inconvenient constraints on the numbers of atoms). More generally, the Ising lattice is potentially a way of modelling continuous systems by discontinuous representations of them, although the difficulty of the algebra in more than two dimensions is an obvious drawback. But should not the model come into its own as a means of dealing with the essentially two-dimensional arrays of atoms in the apparently conducting planes of the new high-temperature superconductors?

That, as it happens, is what U. Löw and V. J. Emery from the Brookhaven National Laboratory, K. Fabricius from the University of Wuppertal and S. A. Kivelson from UCLA have now done (*Phys. Rev. Lett.* **72**, 1918–1921; 1994). The planes containing copper and oxygen atoms in these materials are believed to embody the essence of the super-

conductivity. It is also known that the electronic state of the atoms is not that of electrical neutrality. When electrons are transferred to the electronic conduction band, the atoms left behind are short of electrons and thus behave as electron "holes". Moreover, in the nature of things, these holes are mobile within the copper-oxygen planes.

It may be that what makes high-temperature superconductivity possible is the formation of physical regions within these planes where the density of holes is greater than expected. But how could such a state of affairs come about? To ask that, in a plane in which positive and negative charges are free to move, the positive charges should congregate together, expelling the others, is to require a great deal of the principle that the energy of a static system tends to a minimum. It can only happen if there is some mechanism that yields an energetic benefit when electron holes are close together.

That is the inspiration for a variation on the Ising theme that, in the short run, will probably do more to command the attention of Isingologists than to throw light on the real mechanism in high-temperature superconductors. The novelty of this version of the model is that it takes account of a short-range interaction between the nearest-neighbour vertices that tends to increase the likelihood that neighbouring vertices will be occupied by holes (and *vice versa*) and a long-range interaction between all vertices which is essentially the Coulomb interaction (like charges repel each other and all that).

Solving this model exactly is a lost cause. What Löw and his colleagues have done is to represent the state of each vertex of their square lattice by a three-valued variable which can take the values +1, 0 and -1, representing a hole, the average state of a vertex and the opposite of a hole respectively.

What they are able to show is that the least energetic state of such a system can consist of rectangular patches in which each vertex is occupied by a hole; that may mildly delight the superconductivity community. They also show that in a system such as this, with contradictory long and short-range ordering forces, there may be many ordered phases related to each other by exceedingly complicated phase diagrams if the relative strengths of the interactions are chosen appropriately. It would be rash to think that this will clear up the scandal of why it has taken ten years not to find an explanation for such an important phenomenon, but it is at least a new approach.

John Maddox

Andrew Pomiankowski

have two independent origins, one in *P. pustulosus* and one in the Peruvian population of *P. petersi*. Both phylogenetic reconstructions are plausible. Their likelihood of being correct depends on whether character gains and losses are equally common (supports sensory exploitation) or gains are rare events (supports coevolution).

The newer and more accurate DNA-based phylogenies have weakened the case for pre-existing sensory bias directing the evolution of male sexual traits. This does not mean that sensory systems are

not biased or have no importance in the evolution of signals. But clearly sensory bias is only part of the story, and a full account of preference evolution must take account of the selective consequences of mate choice as well⁸. The hypothesis has nonetheless shown how phylogenetic reconstructions are essential in testing evolutionary ideas about sexual selection. □

Andrew Pomiankowski is in the Department of Genetics and Biometry, University College London, 4 Stephenson Way, London NW1 4HE, UK.

CANCER VACCINATION

Politically incorrect viruses

Peter Parham

THE cure and prevention of cancer by vaccination are the sustaining dreams of tumour immunology. The principle seems simple enough, and almost anything can be construed from mouse models, but the practice has yet to work in humans. Tumours caused by viruses are estimated to affect the lives of over five million people, and for vaccination may offer the best chance for success because their antigens are both singularly foreign and relatively easy to define. It was this premise that underlay a meeting last month*.

The spotlight fell upon four tumour viruses, the diseases they cause and the populations they affect. Vaccination is employed already against one of them, hepatitis B virus, HBV (D. Milich, Scripps Research Institute, California). For the other three — human papilloma virus, human T lymphocyte virus 1 and Epstein-Barr virus (EBV) — vaccine development is still a gleam in the eye. One message to emerge from the meeting was the reaffirmation of the strong ties between the immunology of tumours and of organ transplantation. Today's procedures of clinical transplantation produce their own crop of tumours and opportunistic infections (D. Moss, Queensland Institute of Medical Research, Brisbane) and *in vitro* vaccination procedures designed to prevent the latter infections could well be useful against virally induced tumours (P. Greenberg, University of Washington, Seattle).

Almost by definition, the tumours that trouble humans are those against which there is inadequate immune response. So, from the outset, the battle to press mechanisms of host immunity into the war on cancer has been fought uphill. Although the struggle has led to knowledge, including discovery of the major histocompatibility complex (MHC),

angles on disease have been hard won and their successes pale beside those achieved by vaccination against infectious disease.

The benefit of vaccination lies in the slovenly nature of the immune system. For unknown reasons — but probably economy — the immune system is so slow to get going that its first shots at any new antigen are desultory. Such lethargy can prove deadly. The purpose of vaccination is rehearsal, warming up the immune system under less threatening circumstances in preparation for more serious business ahead.

Obviously the best thing to do to a tumour cell is kill it, and the cytolytic CD8 T cell appears to be the best lymphocyte for the job (C. Melief, University Hospital, Leiden). In designing strategies for vaccination it helps to know the tumour-specific antigens against which T cells can respond, and this is where virally induced tumours may be more tractable than other types of cancer (I. Frazer, Princess Alex-

andra Hospital, Brisbane). The reason is that they express viral proteins that are foreign to the host's immune system and that should be readily recognized by CD8 T cells; furthermore, studies on many virus types have revealed a general mechanism by which CD8 T cells recognize peptides of 8–10 amino acids presented by class I MHC molecules (P. Cresswell, Yale University).

Only a small proportion of people infected by any one of the four viruses goes on to get cancer, and from what is known of these viruses' life cycles there is little reason to believe that they have been selected for their power to produce tumours (P. Doherty, University of Tennessee, Memphis). Doherty proposed that their propensity to cause cancer is because they replicate in proliferating host cells, and are therefore more likely to perturb normal controls on cell division. To illustrate the point he contrasted two herpes viruses — EBV, which is tumorigenic and replicates in dividing epithelial cells, and herpes simplex, which replicates in non-dividing neurons and does not cause tumours. So causing cancer seems to be a secondary consequence of the viral life cycle, and one that can be dramatically affected by other genetic and environmental factors.

Human populations are not equally at risk for the tumours caused by the four viruses; indeed, in selecting their victims the viruses could be accused of political incorrectness. Infection with type 16 papilloma virus can lead to cancer of the cervix in women, but no equivalently devastating outcome in men (M. Stanley, University of Cambridge; P. Beverley, University College and Middlesex School of Medicine, London). Human T-cell leukaemia has historically been associated with the Japanese population (G. de Thé, Institut Pasteur, Paris), and tumours due to HBV are, in number, a problem of the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

False-colour transmission electron micrograph of Epstein-Barr virus particles. This was the first virus to be implicated in human cancer, through its association with Burkitt's lymphoma. Magnification about 1.75×10^5 .

*Vaccines Against Virally Induced Cancers, Ciba Foundation, London, 14–17 March 1994.

developing countries (Milich). But it is EBV that has the most diverse effects. In parts of Africa and Papua New Guinea with endemic malaria, it produces neck tumours (Burkitt's lymphoma) in children, whereas in southern China the prognosis is for nasopharyngeal carcinoma in adults. Within the 'Western-style' societies of temperate climates, childhood exposure to the virus produces cold-like symptoms which are rapidly resolved. In contrast, children esconced in front of the television and protected from infection until late adolescence may contract mononucleosis — the kissing disease of the American campus — a more severe, protracted disease characterized by polyclonal B cell proliferation (Moss).

What looks like a related condition, immunoblastic lymphoma, emerges in transplant patients whose immune systems are deliberately weakened by drugs. This disease may arise either from infection of an EBV-negative recipient by an EBV-positive transplant or to reactivation of latent virus in patients exposed previously to EBV. The former is a hazard in liver transplantation in which young EBV-negative patients often receive tissue from adults (Moss).

Cytomegalovirus (CMV), another herpes virus, does not cause cancer but is the most common opportunistic infection in patients recovering from bone-marrow transplantation. About 25 per cent of such patients succumb to the virus, apparently because they cannot generate CMV-specific CD8 T cells quickly enough from their newly acquired immune system (all immune cells develop from the stem cells of the bone marrow). To plug the gap, Greenberg and co-workers have stimulated and expanded *in vitro* clones of mature CMV-specific T cells from the blood of the donor, and given them to the immunocompromised recipient along with the bone-marrow transplant. This treatment looks promising, in that none of the 12 patients receiving treatment so far has suffered from CMV disease.

Greenberg's strategy could potentially be applied against many types of cancer, particularly those induced by viruses. But for human papilloma virus this approach is stymied by the inability of investigators to generate CD8 T cells with specificity for the virus from the tissues or blood of patients with cervical cancer (Beverly). This state may arise because the epithelial tumour cells are poor at stimulating T-cell responses because they lack costimulatory molecules that are characteristic of professional antigen-presenting cells (J. Allison, University of California, Berkeley). Other findings, however, imply that the papilloma virus is invisible to the immune system (Stanley), which also seems to be the case for EBV in Burkitt's lymphoma (a tumour of the B lymphocyte, a professional antigen-presenting cell) (Moss).

To avoid destruction of virus-infected cells by CD8 T cells, viruses often evolve ways of interfering with the MHC class I pathway of antigen presentation, and this may be the case in cells infected with papilloma virus and EBV. In Burkitt's lymphoma, cellular expression of the transporter proteins that bring viral peptides into close proximity to class I molecules is highly reduced, perhaps explaining why EBNA 1 — the main EBV protein expressed by Burkitt's lymphoma cells — is not an item for immune attack. Indeed, Moss was pessimistic about the prospects of developing a vaccine against Burkitt's lymphoma but more sanguine about those for a vaccine against immunoblastic lymphomas, which express and present a greater range of EBV antigens.

In contrast to the speculative nature of vaccines against EBV and papilloma virus, a vaccine against HBV is already being used in the field and consists of subviral particles of recombinant proteins (Milich). Most HBV-infected individuals resolve the infection and maintain immunity, but a few develop a chronic infection which predisposes to hepatocellular carcinoma. Tumours develop only after 20–30 years of chronic infection, the result of an

accumulation of insults arising from the cycles of lymphocyte-mediated destruction of virus-infected hepatocytes, followed by liver regeneration. More than 250 million people are estimated to be chronically infected with HBV, most of them being in developing countries. In Taiwan and Singapore, vaccination is currently targeted at neonates, who are particularly susceptible to hepatocellular carcinoma after becoming chronically infected by virus from their mothers. But although vaccination may be effective in alleviating infection, its effect on progression to cancer will only be known 50 years from now.

Vaccines against virally induced cancers could help segments of the human population for whom suffering is no stranger. This, however, is not the most attractive market for the world of commerce, which is unlikely to underwrite the inevitably expensive and long-term clinical trials. It is perhaps at that stage that the support of governments will be most urgently needed. □

Peter Parham is in the Departments of Cell Biology, and Microbiology and Immunology, Stanford University, Stanford, California 94305, USA.

PLASMA PHYSICS

Acceleration on a bench top

Robert Bingham

WITH the demise of the Superconducting Super Collider, the question of whether experimental high-energy physics can sustain its vitality has become an urgent one. Foreseeing the possibility of such a crisis, the high-energy physics community began investing in its own future in the form of advanced accelerator research some ten years ago. On page 527 of this issue, Joshi and colleagues report¹ a development that promises an entirely new type of technology for building high-energy accelerators. Their experiment demonstrates electron acceleration in a plasma beat-wave accelerator, which is a new breed of compact accelerators — dubbed 'bench-top' accelerators, although the bench in question would have to be a large one. The results have surpassed all expectations.

Substantial electron acceleration was observed in early experiments on inertial laser fusion and attributed to the production of electron plasma waves moving at high phase velocity. There, the energetic particles are a nuisance — they preheat the central core and prevent compression — but it was these early experiments that led Dawson and co-workers^{2,3} to propose the 'laser wakefield' and 'laser beat-wave' schemes for plasma acceleration of parti-

cles. The latest work¹ uses the second of these schemes.

In both of these schemes, particles are accelerated by high-phase-velocity electron plasma waves generated in the plasma by lasers (see box for details). Electron plasma waves are longitudinal waves with their electric field parallel to the direction of propagation. Electrons moving close to the phase velocity of the plasma wave, which can approach the speed of light, are accelerated by the wave's electric field until they eventually out-run it, rather like a surfer on an ocean wave.

The essential points of the new experiments are as follows. A two-frequency carbon dioxide laser and a 2 MeV electron beam are focused to the same point in a hydrogen plasma at a density of about 10^{16} H^+ ions per cubic centimetre. The intensity of the laser is a few times 10^{14} W cm^{-2} and is roughly constant over a Rayleigh length of about 1 cm. The measured amplitude of the relativistic plasma wave is 30 per cent of its wavebreaking limit, agreeing with theoretical estimates, and giving a theoretical maximum electron energy of about 30 MeV, which is in excellent agreement with observation. The acceleration from 2 MeV to a maximum energy of 30 MeV in about 1 cm

corresponds to an accelerating electric field of 2.8 GV m^{-1} , which is the largest coherent man-made accelerating field yet produced, and 30 times larger than the limit imposed by radiofrequency breakdown in conventional accelerating structures. If the beat wave could be 10 cm long rather than 1 cm as in the experiment, electrons would reach an energy of 300 MeV.

Even if such 'bench-top' machines prove unable to compete with the present high-energy colliders, they could proliferate in research laboratories just as computers have with the advent of PCs and mini-workstations. Apart from their compactness, these plasma accelerators have the interesting feature that they produce a series of very short electron bunches separated by a plasma wave period — here, 1 picosecond (10^{-12} s). These periodic electron pulses could be used to make tiny stroboscopic X-ray or light sources, opening up the prospect of making slow-motion movies of chemical reactions or dynamical biological processes that have never been observed before. In addition, the tiny bursts of X-rays might permit better imaging in biology and medicine using only a fraction of the usual radiation.

In obtaining their results, Joshi and colleagues have overcome some tough practical problems. Their plasma is uniform in density to within 1 per cent, a remarkable value achieved by tunnel ionization of a static fill of hydrogen gas. Essentially the kinetic energy of the electrons increases because of the electric field of the laser — they oscillate in the laser's field. When the laser intensity reaches a critical threshold value in the gas, electrons bound to the nucleus tunnel through the Coulomb barrier, becoming free. The ionization time is of the order of the laser period (10^{-15} s^{-1}) which is extremely small.

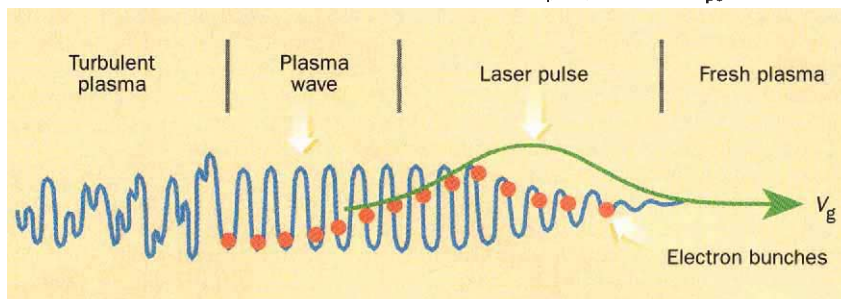
The tendency for strong plasma turbulence to develop, destroying the coherency of the accelerating fields, was overcome by lowering the plasma density, which effectively increased the growth time of plasma instabilities. With the laser pulse and electrons moving close to the speed of light through the plasma, turbulence effects do not have time to establish before the pulse and electrons have passed on.

The work on plasma-based accelerators represents but one area that is being explored by researchers in the advanced accelerator field. Other schemes being investigated at present for high-gradient acceleration are the inverse Cerenkov effect and the inverse free-electron laser effect. Still other researchers, realizing that the next collider will almost certainly be a linear electron-positron collider, are proposing a novel way of building such a device known as a two-beam accelerator.

Laser acceleration of particles

In laser beat-wave experiments, as carried out by Joshi and colleagues, two laser frequencies are incident on a hydrogen plasma producing regions of constructive and destructive interference. The result is a series of light pulses

velocities are close to the speed of light c in a vacuum: $v_p = v_g = c(1 - \omega_{pe}^2/\omega^2)^{1/2}$. Because ω is close to ω' and much larger than ω_{pe} , the Lorentz factor γ_p associated with the beat waves is $[1 - v_p^2/c^2]^{-1/2} = \omega/\omega_{pe} \gg 1$.



moving through the plasma at the group velocity of light.

The plasma electrons feel periodic electromagnetic pressure forces at a frequency corresponding to the difference of the two laser frequencies. If this difference frequency equals the natural oscillation frequency of the electron plasma, ω_{pe} ($= 5.64 \times 10^4 n_e^{1/2} \text{ rad s}^{-1}$, where n_e is the electron plasma density in units of electrons per cubic centimetre) the plasma responds resonantly to these forces and large-amplitude plasma oscillations build up, as shown schematically above.

The laser wakefield scheme is rather different. There, a single laser pulse less than half the length of an electron plasma wave leaves a wake of relativistic electron plasma oscillations.

In both schemes, the electron plasma waves have a phase velocity, v_p , which is equal to the group velocity of light v_g in the plasma. For laser frequencies ω and ω' that are much larger than the plasma frequency, which is the case in Joshi and colleagues' experiment, these

The longitudinal electric field amplitude of these relativistic plasma waves can be extremely large, with a theoretical maximum given by $E = n_e^{1/2} \text{ V cm}^{-1}$, for plasma densities of the order of 10^{18} cm^{-3} , the field strength can be 10^9 V cm^{-1} , equivalent to fields in atoms.

In practice this maximum wave electric field value, set by wave-breaking, is never achieved because the relativistic mass increase of the oscillating electrons changes the plasma frequency and resonance is lost. So the maximum energy to which an electron can be accelerated is $W = 2 \epsilon \gamma_p^2 m_e c^2$ where ϵ is the ratio of the plasma wave amplitude to its theoretical maximum. As an example, if we take a plasma density of 10^{19} cm^{-3} and $\epsilon = 0.3$, 30 per cent of the maximum value, results in an electric field of about 10^9 V cm^{-1} — capable of producing a GeV electron in a distance of 1 cm. Joshi and colleagues no doubt have their sights set on this next step, although higher-frequency lasers than CO_2 will have to be used in these denser plasmas. R. B.

And there are many groups developing an entirely new type of electron lens using focusing by a plasma to increase the luminosity of future linear colliders⁴.

This plays on the fact that relativistic electron beams can be focused by a plasma if the collisionless skin depth c/ω_{pe} is larger than the beam radius. Generally, when a relativistic electron beam enters a plasma, the plasma electrons move to neutralize the charge in the beam on a $1/\omega_{pe}$ timescale. However, if the collisionless skin depth is larger than the beam radius, the axial return current flows in the plasma on the outside of the electron beam and the beam current is not fully neutralized, leading to the generation of an azimuthal magnetic field. Consequently this self-generated magnetic field pinches or focuses the beam in the radial direction. This type of lens exceeds con-

ventional lenses by several orders of magnitude in focusing gradient.

For the beat-wave scheme, the next milestone to be achieved is the GeV energy level, and already Joshi and his team are planning this next step. It is indeed gratifying to see that some of the ideas on alternative acceleration schemes proposed more than a decade ago are coming to fruition. □

Robert Bingham is at the Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK.

1. Everett, M., Lal, A., Gordon, D., Clayton, C. E., Marsh, K. A. & Joshi, C. *Nature* **368**, 527–529 (1994).
2. Tajima, T. & Dawson, J. M. *Phys. Rev. Lett.* **43**, 267–270 (1979).
3. Katsouleas, T. et al. *Proc. AIP Conf. No. 130 Laser Accelerations of Particles* (eds Joshi, C. & Katsouleas, T.) 663–698 (American Institute of Physics, 1985).
4. Chen, P. *Particle Accel.* **20**, 171 (1987).

Improbable views

Andrew Blake

THE problem of shape-from-shading analysis — estimating the relief of a visible surface from its pattern of light and shade in an image — has vexed researchers in vision for around 20 years¹. From an experimental point of view, psychologists seem to disagree whether the human visual system can² or cannot³ make effective use of shading information to estimate even the orientation of a surface, let alone its entire shape. Horn's⁴ elegant demonstration of the reconstruction of a human face by growing height contours outwards from shading highlights showed that inferring shape from shading is theoretically possible, at least under certain conditions. The difficulty comes when, as is usually the case, some parameters of the physical environment are unknown, for example the directions and strengths of the sources of illumination. The parameters of a single light-source can be recovered if certain fairly strong assumptions (isotropy⁵ or smoothness⁶) are made about the shape of the surface, but William Freeman (page 542 of this issue⁷) has now devised a theory founded on the far more general assumption of 'generic illumination'.

This assumption is akin to the concept of 'generic viewpoint' (Fig. 1), already familiar from computational and psychological theories of vision. A viewpoint is generic if it avoids accidental alignments, in the image, of features on a visible object. The assumption of generic viewpoint seems to underlie human visual analysis, explaining why certain line-drawings, for example the well-known Penrose triangle, are perceived as paradoxical⁸ or 'impossible'. The assumption

is so strong that, even for a solid object, if seen from a certain, special viewpoint, an 'impossible' object may be perceived in preference to the real one⁹. Early theories of generic viewpoint applied only to line-drawings of polyhedra; more recently they have been generalized to include drawings of curved surfaces¹⁰.

The assumption of generic illumination can be regarded as the dual of generic viewpoint. The light source and the viewer are dual concepts, as the Table illustrates, each being defined in terms of a cone of light-rays. By analogy with generic viewpoint, generic illumination is the assumption that there are no accidental alignments of object features along illuminant rays (Fig. 2). In principle, this assumption can be invoked to rule out certain interpretations of a shaded image which may be physically consistent but are nonetheless highly improbable. Even so, the generic illumination assumption is not quite in the form that Freeman requires for his theory.

In the case of generic viewpoint, it is known that a 'soft' form of the assumption¹¹ can be obtained to define, for all views and not just one pathological view, the degree of genericity of a given viewpoint. This is a numerical measure that is greater for more probable views but approaches zero as the viewpoint approaches an accidental alignment. In a measure of probability, it can be used to evaluate competing hypotheses for the

shape of a particular visible surface. What Freeman has achieved, this time for generic illumination, is an analogous softening of his assumption, again leading to a measure of probability for competing hypotheses of surface shape. Now the generic illumination assumption is applicable to shaded images rather than simply to line drawings, and this is beautifully illustrated in Fig. 1 of the paper (page 542).

But to what extent does this achievement prefigure a powerful methodology

Viewer-illuminant duality	
Viewer	Illuminant
Viewer position	Light-source position
Visible surface	Illuminated surface
Silhouette	Cast shadow
Blurred image outline	Penumbra

for vision theories, as Freeman believes it does? Although it is true generally that probabilistic models of visual processing are increasingly influential, their application in his paper is somewhat limited in scope.

First, the approach has been shown to work only for the simplest of physical models of reflectance, a single point light-source without shadows or surface patterning. This is a limitation common to many theories of shape from shading. Second, although construction of the theory is based on generic illumination, free of specific assumptions about surface shape, they may emerge nonetheless. Probabilistic measures of generality of illumination like Freeman's can depend, after all, on surface smoothness, having large values for surfaces that are close to being planar or cylindrical. Indeed it does

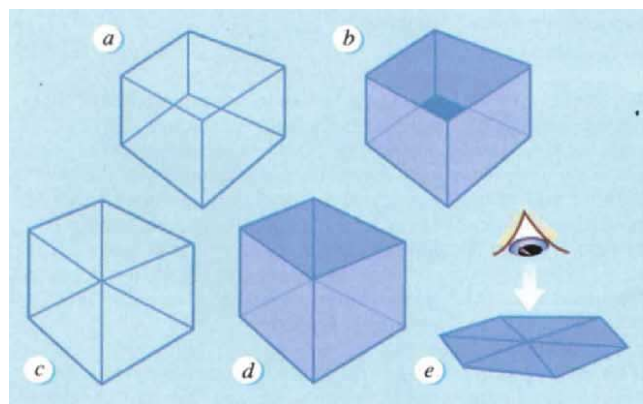


FIG. 1 The assumption of generic viewpoint. The drawing depicted in *a* may be interpreted as a line drawing of a three-dimensional object such as *b*. The drawing *c* could be a line-drawing of an object such as *d*. This interpretation implies, however, that the viewpoint is a special one in which two vertices have coincided in the image. A more plausible interpretation of *c* would therefore be one that involves no accidental alignment, for instance a planar shape such as that shown by *e*.

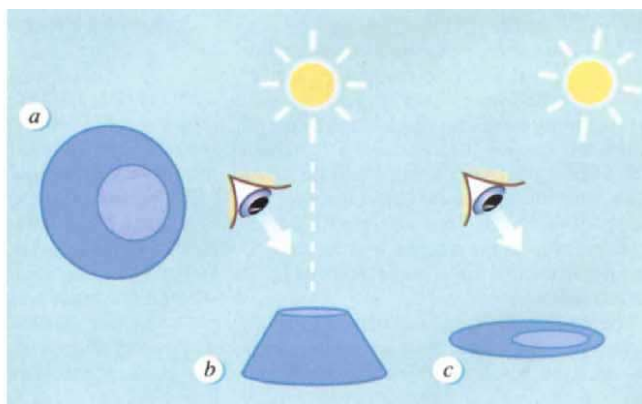


FIG. 2 Generic illumination. Just as the viewer can be in a special position with respect to an object, so can the source of illumination. The image *a* consists of two regions of constant-intensity shading. This could have arisen as the image of a cone *b* but only if the light-source is in a special position, on the axis of the cone. Any perturbation of the light-source's position away from the cone axis will lead to shaded regions whose intensity is no longer constant. A more likely explanation is that *a* is the image of a planar pattern such as *c*.

appear in Freeman's Fig. 1 (though not obviously in his Fig. 3) that the less plausible surface shapes are also less smooth — a greater area of the planar background is disrupted.

One explanation of the link between surface smoothness and generality of illumination is as follows. For diffuse ('Lambertian') illumination, image-intensity is given by the 'rendering function' $I \cdot n$ where n is the surface normal vector at each image position and vector I represents the strength and direction of the light source. In that case Freeman's measure of genericity of illumination (equation 7) is $1/\sqrt{\det(\langle nn^T \rangle)}$, where $\langle \dots \rangle$ denotes spatial averaging over the image. This measure becomes unboundedly large as the visible surface approaches planarity because the normal vectors n span a space of only one dimension, so the $\det(\dots)$ term approaches zero. (Something similar also happens as the surface becomes close to cylindrical.) After all, therefore, as the underlying surface becomes smooth in the sense of approaching planarity, the measure of generality of illumination becomes large. The question is whether this link suggests that Freeman's approach might somehow be limited to smooth surfaces.

Overall, however, Freeman's specific achievement is impressive, both in seeing a way to express probabilistically the general illumination assumption and to some extent verifying experimentally its power to reject incorrect hypotheses about surface shape. The claim for greater generality is not altogether convincing as yet. It will be interesting to see whether computational experiments can establish more precisely the breadth of conditions over which the theory is viable. □

Andrew Blake is in the Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.

1. Horn, B. K. P. & Brooks, M. J. (eds) *Shape from Shading* (MIT Press, Cambridge, Mass., 1989).
2. Bulthoff, H. H. & Mallot, H. A. *J. opt. Soc. Am. A*, **5**, 1749–1758 (1988).
3. Mingolla, E. & Todd, J. T. *Biol. Cybern.* **53**, 137–151 (1986).
4. Horn, B. K. P. *Robot Vision* (McGraw-Hill, New York, 1986).
5. Pentland, A. P. *J. opt. Soc. Am.* **72**, 448–455 (1982).
6. Horn, B. K. P. & Brooks, M. J. *Comp. Vis. Graph. Im. Processing* **33**, 174–208 (1986).
7. Freeman, W. T. *Nature* **368**, 542–545 (1994).
8. Sugihara, K. *Machine Interpretation of Line Drawings* (MIT Press, Cambridge, Mass., 1986).
9. Gregory, R. L. *The Intelligent Eye* (World University, London, 1970).
10. Malik, J. *Int'l J. Comp. Vis.* **1**, 73–103 (1987).
11. Witkin, A. P. *Artif. Intel.* **17**, 17–45 (1981).

GEOCHEMISTRY

Catalysing methane production

Everett L. Shock

THE reactions that produce methane and other light hydrocarbons in sedimentary basins are poorly understood, despite the considerable faith of petroleum geologists in thermal decomposition of organic matter with progressive burial. As with many widely held beliefs (especially in organic geochemistry), the faith of the followers has little to do with the quantitative strength of the hypothesis. Given the geological, geochemical, experimental and theoretical assault now being made on this conventional wisdom, it is not surprising that researchers are proposing new mechanisms and pathways to account for both the widespread occurrence of natural gas and the predominance of methane in that gas. Catalysis is the key, and catalysis by transition metals in organic-rich source rocks was proposed two years ago by Mango¹. Together with Hightower and James, Mango now provides experimental support for this idea on page 536 of this issue².

Mounting evidence indicates that the formation of light hydrocarbons in sedimentary basins is not governed by the thermal instability of organic compounds. Mango has argued for some time^{3–6} that the distribution of light hydrocarbons in petroleum is inconsistent with thermal fragmentation of larger organic com-

pounds. His observation complements the work of L. C. Price⁷, who has assembled extensive geological and geochemical evidence that hydrocarbons survive conditions at which they were once thought to be destroyed.

Meanwhile, thermodynamic calculations led Sato to the conclusion⁸ that increased temperature during burial of organic matter in sedimentary basins does not affect the position of equilibrium in the C–H–O system, but simply increases the rates of otherwise extremely slow reactions. Such reactions lead inevitably to a lower overall free energy for the system; Helgeson and co-workers, including myself, have demonstrated this, and argue for the development of metastable equilibrium states among petroleum, oil field brines and the minerals in reservoir rocks⁹.

The arrival at metastable equilibrium states is apparently driven by the irreversible hydrolytic disproportionation of petroleum to light hydrocarbons, especially methane, and the extent to which the gas can escape from the geological system. Therefore, the mechanisms through which the thermodynamically favoured but kinetically inhibited formation of methane can be catalysed are central to the next steps in understanding geo-

Lightning shot

FASTER than a speeding bullet? A group at Sandia National Laboratories has accelerated a 6-mm-diameter disk of metal to a new record velocity for a macroscopic object. Their multi-stage accelerator works as follows: a powder gun fires a piston down the first, wide-bore barrel to compress a column of hydrogen gas. The gas bursts into a second, narrower barrel where it accelerates a soft-nosed cylindrical projectile to strike the final titanium-alloy disk, sending it flying at velocities up to 15.8 km s^{-1} . This is over 25 times faster than a rifle bullet and, more to the point, about twice the velocity of a satellite in low Earth orbit. So the developers, P. L. Stanton, L. C. Chhabildas and colleagues, are using their 'hypervelocity' gun to reproduce in the laboratory the possible effects of head-on collisions between the proposed space station and orbiting debris.

Proactive profilin

PROFILIN is a protein that keeps cytoplasmic actin in order by sequestering, as was supposed, the monomeric form and thus limiting the proportion that could polymerize. The truth, it turns out, is more complex — and more interesting. Writing in the *Proceedings of the National Academy of Sciences* (**91**, 1510–1514; 1994), T. Finkel and colleagues describe how they have discovered that when a cell line is made to over-express profilin the amount of filamentous actin actually increases. At the same time, the actin bundles that traverse the cytoplasm, and should be the stablest of all, vanish and give place to an increased mass of anisotropic actin close to the plasma membrane. Such effects are also seen when cells are activated by growth factors. How a monomer-binding protein can promote polymerization and how the attendant consequences come about are now matters of some moment.

Low down

OZONE values fell to unprecedented lows over the United States last year, continuing the trend that began in May 1992, says a new report by W. D. Komhyr and co-workers (*Geophys. Res. Lett.* **21**, 201–204; 1994). Their results come from Dobson spectrophotometers at eight ground-based stations on the US mainland, Hawaii and Samoa, five of which have continuous records dating back to the 1960s. For over 40 per cent of the time between January and August 1993 the ozone amounts measured at these stations were more than 3 standard deviations lower than the norm, a consistent drop unrivalled in previous records. Perhaps most worryingly, average values were 8.5 per cent down during May to August, the months when solar ultraviolet insolation is at its strongest.

Loosened bonds

PLASTIC packaging is splendidly hygienic for the food manufacturer, but a major nuisance as garbage. Biodegradable polymers, claimed to rot rapidly when thrown away, haven't solved the problem. Daedalus has a new suggestion.

He points out that the long-chain molecules of plastic materials are made by linking many monomer units together. Many such chains are quite capable of reversing this reaction, and coming apart again. The acetal resins and some nylons, for example, are liable to 'unzip' by repeated loss of the end monomer unit on the chain. Their molecules have to be stabilized by attaching a stable 'cap' on each end, to prevent the unzipping reaction from starting.

Daedalus is now designing a plastic film whose polymeric chains do not lie flat within it, but widthways across it, running from one face to the other. The end-groups on one face are firmly capped with a stable end-group. Those on the other face are more precariously capped with an end-group that is easily hydrolysed away by water, or oxidized by air. Formed into a two-layer 'sandwich', with the stable sides facing outwards and the unstable ones pressed together in the middle, such a film would be quite permanent. It could be used for bottles, bags, sachets and packaging of all kinds. But the act of opening the package would tear the film. Air or water would get at the vulnerable centre of the sandwich. The whole thing would then unzip rapidly back to monomer.

DREADCO's chemists are now trying to realize this vision. They are exploiting the known technique of electro-initiated polymerization. A polymer is grown on an electrode immersed in monomer solution; the applied voltage maintains a reactive charge on the end-group of each growing chain. On a flat electrode the parallel polymer chains should extend, like a fully dense microscopic growth of hair, into a sort of polymeric Langmuir-Blodgett film. When it is thick enough, the polymeric film will be removed from the solution, end-capped on the exposed face, stripped off the electrode, folded into sandwich form, and sealed. The process could even be run continuously on a rotating cylindrical electrode.

Autodestructive film will transform the packaging market. Packages will last exactly as long as required, and will then degrade automatically. A gaseous monomer would simply evaporate. But Daedalus would prefer a combustible liquid monomer which could be reclaimed later. Consumers could dump their old packaging in a barrel, and drain off the accumulating monomer as free motor fuel.

David Jones

chemical transformations of organic compounds in the crust of the Earth.

Transition metals are the catalysts that Mango and co-workers are pursuing. They find that organic-rich source rocks, which are heat-treated to eliminate their inherent petroleum generation potential, have the catalytic ability to generate gas from hydrogen and *n*-alkenes. The gas produced has a composition strikingly similar to that of natural gas, methane being the predominant product (about 90 mol per cent). Their result contrasts with other experimental attempts to produce natural gas from organic matter at high temperatures in which the relative yields of methane are considerably lower. The implication is that something in the inorganic make-up of the source rock is the catalyst for the reaction, and Mango and co-workers suggest that it may be the unusually high concentrations of nickel and/or vanadium.

Although the actual catalytic culprit is not known, the results are consistent with the transition-metal catalysis scheme that Mango proposed¹ in 1992. It seems plausible that the catalytic process involves an oxidation-reduction couple, as something must be oxidized when carbon (in whatever form) is reduced to methane. Transition metals are one possibility, and the source of hydrogen in the natural setting is either the organic matter itself or reactions between minerals and water. So it is especially intriguing that Mango and co-workers found an increase in catalytic activity on adding a small amount of water to their experiments. As they note, however, there is apparently an effect of the concentration of H₂O on the results of the reaction.

The success of this bulk-rock study opens the door for controlled experiments to characterize the catalytic properties of the source rocks' various inorganic constituents: for instance, transition metal complexes in carbonaceous material or the structure of mineral surfaces. Mango and co-workers have not characterized in detail the mineralogy, petrology, texture or geochemical nature of the source rock before or after their laboratory work. Subsequent studies could be devised to investigate the likelihood that coupled organic and mineral transformation reactions provide the kick for methane generation.

On the other hand, the conditions imposed in the laboratory rarely include all the possibilities of the natural system. Some microbiologists^{10,11} seem to be comfortable with the idea that heat-loving microorganisms may thrive at temperatures as high as 150 °C. If so, then the biosphere could extend throughout the temperature range of the traditional 'oil window' in sedimentary basins. This raises the possibility⁹ that the metabolic processes of heat-and-pressure-loving organisms

(also known as hyperthermobarophiles) are the methane-forming catalysts in sedimentary rocks.

Current evidence cannot rule this idea out; indeed, the observation of anomalously high concentrations of methane in submarine hydrothermal vent fluids from the Juan de Fuca ridge¹² may provide some clues. Thermodynamic calculations¹² demonstrate that the concentration of methane is above that set by equilibrium involving H₂, H₂O and CO₂ at the temperatures of the vent fluids (317–368 °C). However, the concentrations of all four gases are consistent with equilibrium at about 275 °C. A catalysed approach towards equilibrium at any lower temperature, say 150 °C, would be consistent with even higher concentrations of methane, which might be diluted during mixing with fluids from other sources to provide the concentrations found in the vent fluids. The possibility of high-temperature methanogenesis is made more intriguing by the observation of particulate DNA in fluid samples from the same vents¹³.

At present, there is a bewildering choice of possible methane-forming catalysts in natural systems, all awaiting further experimental and theoretical investigations. Although still in their infancy, studies which go beyond the phenomenological and address the thermodynamic and kinetic constraints that lead to methane formation will modify thinking about geochemical processes involving fluids, aqueous solutions, gases, organic matter and volatile compounds in general. The applications could span an enormous variety of topics from petroleum geochemistry to sea-floor hydrothermal processes, and from life at high temperature to the study of the evolution of icy satellites in the outer Solar System. □

Everett L. Shock is in the Department of Earth and Planetary Sciences and the McDonnell Center for the Space Sciences, Washington University, St Louis, Missouri 63130, USA.

1. Mango, F. D. *Geochim. cosmochim. Acta* **56**, 553–555 (1992).
2. Mango, F. D., Hightower, J. W. & James, A. T. *Nature* **368**, 536–538 (1994).
3. Mango, F. D. *Science* **237**, 514–517 (1987).
4. Mango, F. D. *Geochim. cosmochim. Acta* **54**, 23–27 (1990).
5. Mango, F. D. *Geochim. cosmochim. Acta* **54**, 1315–1323 (1990).
6. Mango, F. D. *Nature* **352**, 146–148 (1991).
7. Price, L. C. *Geochim. cosmochim. Acta* **57**, 3261–3280 (1993).
8. Sato, M. in *Fluid–Mineral Interactions: A Tribute to H. P. Eugster* (eds Spencer, R. J. & Chou, I.-M.) 271–283 (The Geochemical Society, San Antonio, 1990).
9. Helgeson, H. C., Knox, A. M., Owens, C. E. & Shock, E. L. *Geochim. cosmochim. Acta* **57**, 3295–3339 (1993).
10. Daniel, R. M. *Origins Life Evol. Biosphere* **22**, 33–42 (1992).
11. Segerer, A. H. *et al. Origins Life Evol. Biosphere* **23**, 77–90 (1993).
12. Lilley, M. D. *et al. Nature* **364**, 45–47 (1993).
13. Deming, J. W. & Baross, J. A. *Geochim. cosmochim. Acta* **57**, 3219–3230 (1993).

Dangers of asteroid deflection

SIR—There are thought to be about 1,000 Earth-crossing asteroids, each with a diameter >1.5 km, although only about 50 have so far been identified¹. A collision of one of these asteroids with the Earth would release an energy equivalent to 10^5 MT, disrupt the ecosphere, terminate agriculture, and kill a significant fraction of the world's human population (see, for example, ref. 2). The *a priori* risk of such a catastrophe during the next century is about one in 5,000.

The proposed 'Spaceguard' survey^{2,3} could, in 25 years, find about 95% of these potentially threatening asteroids. We can predict with $>99.9\%$ certainty that no object will be found on a trajectory posing any danger during at least the next century, in which case Spaceguard would seem to reduce the *a priori* risk of a global impact catastrophe by nearly two orders of magnitude. In the improbable case that an asteroid were found on a threatening trajectory, the most likely warning time would be several decades or more, long enough to mitigate the hazard. The most efficient approach with existing technology would be to deflect the object through a herding series of standoff nuclear explosions^{4,5}.

This proposal is a double-edged sword. If we can perturb an asteroid out of impact trajectory, it follows that we can also transform one on a benign trajectory into an Earth-impactor. For example, the asteroid 1991 OA could in 2070 be deflected into Earth-impact trajectory⁵ with an aggregate yield of only about 60 MT. Although a single asteroid can more readily be deflected away from, than into, an impact trajectory, there is not an orders-of-magnitude difference in technical effort; but there are orders-of-magnitude more Earth-crossing asteroids that can be induced to impact the Earth than will do so on their own. With a Spaceguard-like inventory of such asteroids and a launch-ready deflection system of nuclear-armed missiles, it might take only a few years to identify a suitable large asteroid, alter its orbit through a series of nuclear explosions with individual yields of about 10 MT (available in existing arsenals), and send it crashing into Earth⁵. There is no other way known in which a few nuclear weapons could by themselves threaten the global civilization.

In our view, development of this asteroid-deflection technology would be premature. Given twentieth-century history and present global politics, it is hard to imagine guarantees against eventual misuse of an asteroid deflection system commensurate with the dangers such a system poses. Those who argue that it would be prudent to prevent catastrophic impacts with annual probabilities of 10^{-5}

will surely recognize the prudence of preventing more probable catastrophes of comparable magnitude from misuse of a potentially apocalyptic technology.

It is of course sensible to seek cost-effective reduction of risks from all hazards to our civilization — even low-probability hazards, of which many may remain unidentified. At a total cost of some \$300 million, Spaceguard arguably constitutes a reasonable measure of defence against the impact hazard. But premature deployment of any asteroid orbit-modification capability, in the real world and in light of well-established human frailty and fallibility, may introduce a new category of danger that dwarfs that posed by the objects themselves.

Carl Sagan

Laboratory for Planetary Studies,
Cornell University,
Ithaca, New York 14853, USA

Steven J. Ostro

Jet Propulsion Laboratory,
Pasadena, California 91109, USA

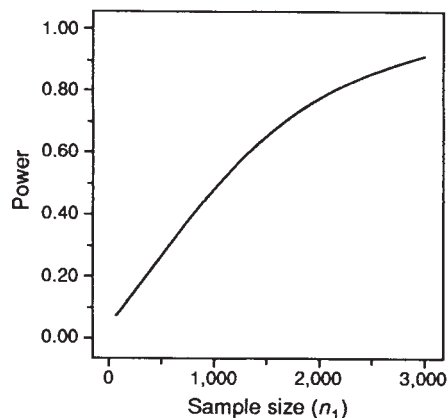
1. Shoemaker, E. M., Wolfe, R. F. & Shoemaker, C. S. in *Global Catastrophes in Earth History* (eds Sharpton, V. L. & Ward, P. D.) 155–170 (Geol. Soc. Am. Pap. 247, 1990).
2. Chapman, C. R. & Morrison, D. *Nature* **367**, 33 (1994).
3. Morrison, D. *Acta astronautica* **30**, 11–16 (1993).
4. Ahrens, T. J. & Harris, A. W. *Nature* **360**, 429–433 (1992).
5. Harris, A. W., Canavan, G. H., Sagan, C. & Ostro, S. J. in *Hazards Due to Comets and Asteroids* (ed. Gehrels, T.) (University of Arizona Press, Tucson, in the press).

Statistical scrotal effect

SIR—Nieschlag *et al.*¹ used historical data to examine the hypothesis that testosterone influences longevity in men. They examined mean life span in 50 castrati, famous as singers during the period 1581–1858, compared to the mean life span of 50 similarly famous 'intact' singers born at the same times. The difference in mean life span (1.2 yr) was not significant in a two-tailed *t*-test ($P=0.65$), leading the authors to conclude: "These data show that prepubertal removal of the testes had no influence on the longevity of men." But the authors have fallen into a very common statistical trap — they conclude that a failure to demonstrate an effect implies that the effect does not exist, that an unrejected null hypothesis is true. A more reasonable conclusion from their data would have been that there is no evidence that the mean differences are not equal, which is not the same thing as concluding that the means are equal.

It is instructive to examine this data set from a power-analysis point of view. Let us assume, as the authors did, that (1) age at death is normally distributed (this is almost certainly untrue); (2) the variances

of age at death in castrati and intact singers are equal and estimable from the sample variances; and (3) the appropriate null hypothesis is that the mean age at death is the same (given the authors' preamble, a one-tailed hypothesis would have been more appropriate). Then, if the real difference in age at death were 1.2 yr, what is the probability (power) of detect-



Effect of sample size on power of a two-tailed *t*-test for independent means to reject a false null hypothesis if the difference in means = 1.2 yr ($\alpha=0.05$, $s=13.95$).

ing such a difference with a sample of size 50 in each group? Using a pooled variance of the ages at death, the probability is only 0.071 (see figure). It is therefore not at all surprising that the authors did not detect a significant effect — they had almost no chance of doing so. If the effect of castration were to change mean lifespan (in either direction) by 1.2 yr, then, given the observed variance structure, the authors would need a sample of size $>2,120$ (of each type of singer) to have an 80% chance of detecting such an effect with a maximal probability of 0.05 of making a type-1 error. Examined in another way — what is the smallest difference in mean age at death that could have been detected with this data set if the maximal probability of a type-1 error was 0.05 and we wanted an 80% chance of detecting such a minimal difference if it existed? The answer is 9.1 yr. If the difference in longevity were that large, we probably wouldn't be engaged in statistical arguments about its validity!

The best way, of course, to conclude that a null hypothesis is probably true is by replication; if repeated attempts to falsify the hypothesis always fail, then one begins to be increasingly convinced that the hypothesis is probably true. It is clearly ill advised to conclude the truth of a hypothesis based on a single sample, in the absence of a consideration of what effect size was thought to be worth detecting, or in the absence of a power analysis of the effect size that was observed. Readers wishing a gentle and informative introduction to power analysis should consult the

wonderful paper by Peterman²; that seminal contribution should have what I term here the "statistical scrotal effect" — it should cool the ardour of most hypothesis testers (testees?).

Stephen M. Smith

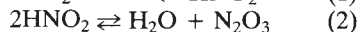
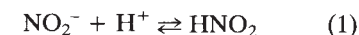
Department of Biology,
University of Waterloo,
Waterloo, Ontario, Canada N2L 3G1

1. Nieschlag, E., Nieschlag, S. & Behre, H. M. *Nature* **366**, 215 (1993).
2. Peterman, R. M. *Can. J. Fish. aquatic Sci.* **47**, 2–15 (1990).

Stomach NO synthesis

SIR — In man, nitrate is concentrated in the saliva and rapidly converted to nitrite by facultative anaerobic bacteria on the surface of the tongue^{1–3}. In the stomach, acidification of this nitrite will result in the

formation of nitrous acid (pK_a 3.2, equation 1) and then nitrogen oxides (equations 2, 3). We propose that this novel mechanism for NO (nitric oxide) generation in the lumen of the stomach is important as a defence against swallowed pathogenic microorganisms.



We tested the saliva of ten people working in our laboratory who had fasted overnight. Salivary nitrite varied from 23 to 220 μM (mean 114) rising to 409–1,890 μM (mean 1,030) 45 min following ingestion of 200 mg potassium nitrate solution. Nitrite solutions generate NO on acidification at a rate dependent on both nitrite and hydrogen-ion concentration; 200 μM nitrite, when acidified to pH 2, results in an NO concentration of approximately 600 nM, several orders of magnitude greater than that required to cause vasodilation⁴.

The yeast *Candida albicans* retains viability when incubated with acid alone for an hour at pH 3 but is destroyed when 250 μM nitrite is added to the incubation medium (*a* in the figure). *Escherichia coli*, which is closely related to *Salmonella* sp., *Shigella* sp. and other pathogenic Enterobacteriaceae, when incubated for 1 h at pH 3 shows sensitivity to as little as 10 μM nitrite (*b* in the figure). As more than 1 litre of saliva is usually swallowed each day, it is likely that microbicidal concentrations of nitrite exist in the stomach, especially in people consuming a high-nitrate diet (mainly provided by green vegetables in developed countries)⁵.

NO, which is generated when nitrite is acidified, readily diffuses through cell membranes and has a high affinity for iron-sulphur-containing respiratory enzymes and damages bacterial DNA⁶. When produced enzymatically by activated leukocytes, nitric oxide will kill various gut pathogens^{7–11}.

Although there has been concern that the enterosalivary circulation of nitrate

may result in the formation of harmful nitrosamines^{12,13}, we suggest that this alternative route of NO synthesis has developed for the specific purpose of prevention of lower gastrointestinal infection.

Although we have considered the effect of acidified nitrite only on *C. albicans* and *E. coli*, this mechanism may also be important in providing protection from other serious gut pathogens which, when swallowed, may cause duodenal ulceration (*Helicobacter pylori*), amoebic dysentery and chronic intestinal parasitism.

Nigel Benjamin, Fionnuala O'Driscoll,

Hamish Dougall, Callum Duncan,

Lorna Smith & Michael Golden

Department of Medicine and Therapeutics,

Hamish McKenzie

Department of Medical Microbiology,

University of Aberdeen

Medical School,

Forresterhill, Aberdeen AB9 2ZD, UK

Viral-induced extinctions unlikely

SIR — Emiliani in Scientific Correspondence¹ implies that the existence of phytoplankton-infecting viruses, including viruses that may constrain the blooms of phytoplankton², provides support for viral-induced extinctive evolution. Emiliani had earlier proposed³ that the sudden extinctions and appearances of marine protists could be due to pathogen (fungal and virus) infestations. But in my view, viral pathogen-induced extinctions seem unlikely.

Viruses are obligate parasites, and their survival depends on the survival of the host. If a virulent virus reduced the population of its host, the virus population would similarly be reduced until a balance (steady state) was achieved. Indeed, the phytoplankton-infecting viruses² reduced the ability of phytoplankton to grow only when the cultures reached stationary phase. A virulent virus can only eliminate a host population if the virus can also live in another host which it does not kill. In the absence of an alternative host population for these viruses it is improbable that lethal mutant viruses arose repeatedly during evolution and completely eliminated their hosts. The finding of an alternative host organism for these viruses would provide more compelling support for Emiliani's theory of extinctive evolution³.

Victor E. Buckwold

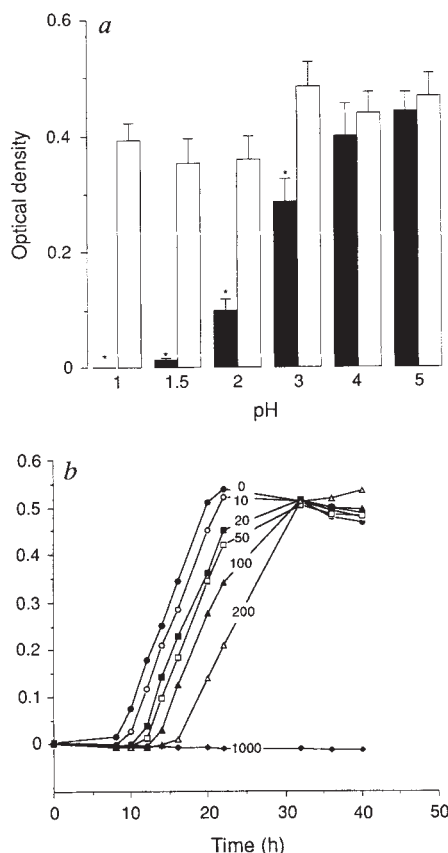
Department of Microbiology,

University of Southern California

School of Medicine,

Los Angeles, California 90033, USA

1. Emiliani, C. *Nature* **366**, 217–218 (1993).
2. Shuttles, C. A., Chan, A. M. & Cottrell, M. T. *Nature* **347**, 467–469 (1990).
3. Emiliani, C. *J. theor. Biol.* **97**, 13–33 (1982).



a, Effect of exposure to nitrite and differing hydrogen ion concentrations on the survival of *C. albicans*. Open bars, growth of *C. albicans*, measured by optical density, in Sabarouds broth, 9 h following exposure to acid alone (phosphate/citrate buffer) for 1 h. Closed bars, growth following exposure to acid and 250 μM sodium nitrite. Asterisk, significant difference from control ($P < 0.05$, Mann-Whitney *U* test, mean of 20 experiments). *b*, Growth curves of *E. coli* in nutrient broth (Oxid CM1) following exposure to phosphate/citrate buffer, pH 3, with increasing concentrations of nitrite (μM) for 1 h. Mean of 16 experiments.

1. Sasaki, T. & Matano, K. *J. Fd Hyg. Soc. Jap.* **20**, 363–369 (1979).
2. Ishiwata, H., Tanimura, A. & Ishidate, M. *J. Fd Hyg. Soc. Jap.* **16**, 89–92 (1975).
3. Tannenbaum, S. R., Weisman, M. & Fett, D. *Fd Cosmet. Tox.* **14**, 549–552 (1976).
4. Palmer, R. M. J., Ferrige, A. G. & Moncada, S. *Nature* **327**, 524–526 (1987).
5. Knight, T. M., Forman, D., Al-Dabbah, S. A. & Doll, R. *Fd Cosmet. Tox.* **25**, 277–285 (1987).
6. Wink, D. A. *et al. Science* **254**, 1001–1003 (1991).
7. Liew, F. Y., Li, Y. & Millott, S. *J. Immun.* **145**, 4306–4310 (1990).
8. Malawista, S. E., Montgomery, R. R. & van Blaricom, G. *J. clin. Invest.* **90**, 631–636 (1992).
9. Green, S. J. *et al. Infect. Immun.* **61**, 689–698 (1992).
10. Denis, M. *J. Leuk. Biol.* **48**, 380–387 (1991).
11. Cenci, E. *et al. Eur. J. Immun.* **23**, 1034–1038 (1993).
12. Tannenbaum, S. R., Sinskey, A. J. & Bishop, W. *J. natn. Cancer Inst.* **53**, 79–84 (1974).
13. Spiegelhalter, B., Eisenbrand, G. & Preussmann, R. *Fd Cosmet. Tox.* **14**, 545–548 (1976).

Seeds of revolution

Valery N. Soyfer

Origin and Geography of Cultivated Plants. By N. I. Vavilov. *Cambridge University Press*: 1992. Pp. 496. £75, \$120.

NIKOLAY Vavilov was one of a small number of scientists who dedicated themselves to the study of the origin of plant species and the geography of cultivated plants. Of Vavilov's 528 publications, 64 were devoted to this topic, and this book includes 24 of the key ones. So it is inaccurate to describe it as "a collection of all of Vavilov's works on the origin and geography of cultivated plant species". It is equally difficult to agree with the editors' claim that "For the first time, all the major works of N. I. Vavilov dealing with the problems of the geography of cultivated plants and the establishment of their centers and origins are assembled in the present collection of papers." Many of these works were published in the West during Vavilov's lifetime, and at least three sets of collected works were published in Russian editions after his posthumous rehabilitation.

Vavilov was born in Moscow on 25 November 1887, the son of a prosperous Russian merchant who became a millionaire. After graduating from the Moscow Agricultural Institute in 1911, Vavilov spent almost a year in the laboratory of William Bateson in England. After the outbreak of the First World War, he returned to Moscow and started his research on the resistance of plants to disease. He then turned to the study of wild relatives of cultivated plants, and formulated the idea that all domesticated plants originated in areas of human activity in prehistoric times. To support his hypothesis, Vavilov organized expeditions to areas where the most ancient human settlements supposedly originated. He identified five major centres of origin of cultivated plants by 1930, and later increased this number to seven.

By 1940, Vavilov and his pupils had collected more than 250,000 samples of seeds from all over the world (or 160,000 samples according to this book). Under his direction, experimental stations (more than 400 throughout the Soviet Union) organized the botanical, taxonomic, cytological and sometimes even genetic investigation of these samples. Thus, the first world bank of plant genes was established.

Indeed, Vavilov played a unique role in the organization of Soviet science. He organized and was the first president of the All-Union Academy of Agricultural Sciences, under the auspices of which he helped to open no fewer than 18 research institutes in the Soviet Union.

He was also an outspoken advocate of

the new socialist order and helped to develop the new socialist science, as the following example shows. The terrible results of the widespread collectivization of Soviet agriculture between 1929 and 1930, when all grain was confiscated by the army and when many traditional Russian grain varieties disappeared, presented the Soviet government with the urgent task of producing new grain varieties. Vavilov believed that samples in the Institute of Plant Industry's collection



Vavilov (left) with William Bateson in 1925.

could help in the selection of new grain varieties, and he was responsible for the ambitious programmes for Soviet selection and plant breeding (several articles in this book reflect these views). However, this task was much more difficult than Vavilov expected because of genetic obstacles. Vavilov was interested in genetics but lacked a deep understanding of the field and believed that plants originating from different parts of the globe could easily be adapted to the Russian environment through crossing with other varieties. He thought that Lysenko's vernalization method would help in implementing this programme.

Vavilov constantly promoted Lysenko between 1929 and 1935 and publicly supported him. But in 1934–35, Lysenko, in his attempts to capture power in Soviet biology and agronomy, accused Vavilov of scientific mistakes, and later qualified

them as political mistakes. He rejected, for example, Vavilov's ideas on the origin of plants and especially his "law of homologous variations". Vavilov thought that closely related species, genera and families have similar variations of characteristics. Although this theory preceded a real understanding of the chemical nature of heredity and the possible evolutionary paths of genetic sequences, it reflected evidence of similarity of changes in genetically closed organisms with definite genetic relationships. Lysenko's struggle against Vavilov caused public debate, and many of Lysenko's followers found it profitable to raise their voices against Vavilov. A graduate student planted in Vavilov's institute even said publicly: "Stalin clearly explained that Lysenko's views are correct and Vavilov's are not required for the building of Commun-

ism." During scientific meetings, Vavilov was forced to listen to accusations of mistakes, of the uselessness of his theory and so on. Blatantly adversarial papers against Vavilov were published by Grigory Shlykov (Vavilov's reply is reprinted in the book), and on a purely political basis, Shlykov and other Lysenko supporters declared him an enemy of the socialist state. He was arrested and imprisoned in Saratov, where he died from starvation on 26 January 1943.

The book was prepared for publication when *glasnost* had been declared, but did not yet exist. As a result, the reader will not find anything about this bitter and vicious treatment of Vavilov by his countrymen. The translator has tried to outline the main events in Vavilov's life in the preface to the book. One is especially interesting: "He [Vavilov] was even made a Foreign Member of the British Royal

Society, a rare honor since only 50 such are selected worldwide". Actually, the story is more complicated. In 1943, British scientists decided to undertake strong measures to support the arrested academic. For a long time, membership in the most renowned society of scientists worldwide, the Royal Society of London, had been offered to only a small coterie of scientists. Now membership was offered to Vavilov. The British Ambassador visited the Soviet Foreign Ministry and asked for help in reaching Vavilov to obtain his signature on a special diploma that would certify his readiness to accept the membership. This request presented a serious dilemma to the Soviet authorities. The solution was in the spirit of Machiavelli. Vavilov had a younger brother, the physicist Sergei Vavilov, who was later appointed president of the Soviet Academy of Sciences by Stalin. Sergei was ordered to sign the diploma instead of his brother, which he did, and the signed document was returned to the British. The signature was recognized as a forgery and attempts at reconciliation were made through diplomatic channels, but they were unsuccessful. The doors of

the prison cell stayed shut.

Although the translation from Russian is accurate, there are several flaws in the book. First, many titles of articles are different from Vavilov's own titles. Second, there are several mistakes in the "Translator's Foreword". For example: "in 1938 Vavilov was ousted from the Presidency of the Academy of the Agricultural Sciences, of which he was the initiator and first president. He was replaced by Lysenko". Actually, Vavilov lost his position as president of the academy in June 1935 and was replaced by an old communist, A. I. Muralov. In 1937 Muralov was arrested, and the interim presidential position was given to Meister, who was arrested in 1938. Only after Meister's arrest was Lysenko appointed president of the academy. The translator writes that Vavilov "personally visited and made extensive research in no less than 52 different countries". My own research has given me a number closer to 30, others suggest 35. □

Valery N. Soyfer is in the Laboratory of Molecular Genetics, George Mason University, Fairfax, Virginia 22030, USA.

Natural truths

David L. Hull

Smoke and Mirrors. By James Robert Brown. Routledge: 1994. Pp. 200. £35 (hbk); £11.99 (pbk).

We Have Never Been Modern. By Bruno Latour. Harvard University Press/Harvester Wheatsheaf: 1993. Pp. 157. \$29.95, £35 (hbk); \$12.95, £13.95 (pbk).

BROWN and Latour purport to be talking about the same subject — science — but anyone reading these two books would be hard pressed to find much in the way of overlap in either their content or the language in which they are couched. Brown is concerned with refuting various arguments presented by analytic philosophers over the past couple of decades against a realistic interpretation of science and suggests somewhat sheepishly a Platonic alternative. He discusses such topics as hypothetico-deductivism, Arthur Fine's natural ontological attitude, Richard Rorty's solidarity version of relativism, Michael Ruse's Darwinian naturalism, constancy of reference across contexts, thought experiments and the structure of data. His exposition is clear and concise, possibly too concise. He assumes that his readers are already familiar with the issues at hand and says just enough to remind them of the general outlines of the problems and past solutions. He also assumes that the reader has a background in physics sufficient to understand, for example, Einstein, Podolsky and Rosen's 1935 thought experiment and Bell-type inequalities.

Analytic philosophers have been analysing science for a long time. Because I was raised in this tradition, the language and sorts of arguments employed sound perfectly natural to me. Brown exhibits both the strengths and weaknesses of the tradition. The greatest strength of analytic philosophy is its clarity. Brown cuts through slag heaps of verbiage to present issues as clearly as anyone could wish. Initially, at least, analytic philosophers increased our understanding of science, but rapidly they turned in on themselves, forgetting science as they followed the snarled skein of logical paradoxes generated by their analyses. The only role that science plays is to provide stock examples to illustrate philosophical points.

Latour's book first appeared in 1991 in French. Although Latour rewrote his text before translation into English, he retained the speculative and "very Gallic" character of his essay. His main concern is to argue against postmodern views of science. Many people, I among them, are growing weary of the mishmash of self-important denunciations and literary allusions termed 'postmodernism' (Emily



THE astronomer, draughtsman and publisher Johannes Hevelius, as depicted in his first book *Selenographia* (1647), a treatise on all the heavenly phenomenon accessible with the telescope. His contribution to astronomy is re-evaluated in *Renaissance and Revolution* edited by J. V. Field and Frank A. J. L. James, a diverse collection of 15 essays on the history of science in Europe between about 1400 and 1750. CUP, £37.50, \$49.95.

Dickinson as a Marxist feminist, for example). Latour is even more exasperated with postmodernism than I am. He cannot find "words ugly enough to designate this intellectual movement", but many of his readers are likely to find it difficult to separate Latour from the postmodern networks in which he has participated for so long. Brown, for one, reads Latour as an anti-realist social constructivist. Facts are "constructed", "fabricated", not to mention "socially negotiated." According to Brown, Latour is confusing changes in belief with changes in truth value. The Earth did not change its shape as people changed their minds about its shape, and just try negotiating the AIDS virus into a benign commensal.

Latour's writing at its clearest is so overwrought that it always leaves plenty of room for alternative interpretations, but in claiming to reject postmodernism, I think Latour is just being ornery, turning on his own people as they ease into comfortable, complacent middle age. In fact, Latour makes short work of the conceit that provides the title of his book. According to Latour, modernity (accent on the second syllable) has never begun. Hence, postmodernism is impossible. How much more ludicrous can one be than to "claim to come after a time that has not even started"? Of course, the existence of modernity depends on how one defines it. According to Latour "modernity has nothing to do with the invention of humanism, with the emergence of the sciences, with the secularization of society, or with the mechanization of the world". Instead, he would have us understand modernity in terms of the introduction of a "total distinction between humans and nonhumans on the one hand and between purification and mediation on the other". For pre-moderns, nature did not exist separate from culture, nor nonhumans from humans, but formed a seamless whole. True moderns tore this fabric asunder.

Latour sees the beginnings of this great schism in the dispute between Thomas Hobbes and Robert Boyle over the air pump. To the extent that there is a modern world, two seventeenth-century Englishmen invented it — quite an admission for someone as Gallic as Latour. Latour proposes to heal the maladies of modernity by recognizing (inventing? constructing?) a single nature-culture populated by quasi-objects which are, at the same time, quasi-subjects. Latour's quasi-objects are "in between and below the two poles [of nature and society], at the very place around which dualism and dialectics had turned endlessly without being able to come to terms with them". They are much more social, fabricated and collective than the "hard" parts of nature, but they are also not mere "arbitrary receptacles" of society. Conversely,

they are "much more real, nonhuman and objective than those shapeless screens on which society . . . needed to be 'projected'." I am afraid Latour's nonmodern view remains as obscure to me now as when I first read his book. Of course, the problem may be that my Latouresque is not much better than my French.

Brown and Latour agree about one thing — the evils of relativism. According to Brown, "Relativism is a repugnant doctrine — reprehensible, repulsive and easily refuted". Latour presents one quick refutation. He opposes social relativists because they are constructivists where nature is concerned and realists with respect to society. But, as Latour insists, "Society, as we now know, is no less constructed than Nature". Brown would hardly be content with Latour's refutation. In response to Rorty's claim that nothing answers to the name of 'objectivity' in science, only solidarity in the sense of unforced agreement with one's peers, Brown objects that no such solidarity exists in science. If scientists do anything, it is disagree with each other. Reflexivity leads Hilary Putnam to ask: "Do the standards of Rorty's [European] culture . . . really require Rorty's 'cultural peers' to assent to what he has written? Fortunately, the answer is negative".

Both Brown and Latour are philosophers writing about philosophical problems for other philosophers while the main character in the drama stands forgotten on an empty stage. Philosophers have the right to argue with each other about philosophy but not to pretend that in doing so they are engaged in the study of science. At least Brown uses examples from science to illustrate his Platonism, while Latour all but ignores science. In Latour's book, there is no history, no people, only ideas fighting with ideas. □

David L. Hull is in the Department of Philosophy, Northwestern University, Evanston, Illinois 60208, USA.

■ Also recently published is *Taking the Naturalistic Turn: How Real Philosophy of Science is Done*, organized and moderated by Werner Callebaut (University of Chicago Press, \$85, £67.95 (hbk), \$29.95, £23.95 (pbk)). The book contains the transcripts of Callebaut's discussions with 24 biologists and psychologists as well as historians, philosophers and sociologists of science. Philosophers of science, he argues, traditionally have ignored the details of scientific research, resulting in theories that lack relevance either to science or to philosophy in general. The interviewees debate the limitations of this tradition and the advantages of the "naturalistic turn" — the idea that matters of fact are as relevant to philosophical theory as they are to science. Among those interviewed is David L. Hull.

Exposed to the elements

Felix Franks

Air and Water: The Biological and Physics of Life's Media. By Mark W. Denny. Princeton University Press: 1993. Pp. 341. \$39.50, £24.95.

SELDOM does one come across a science book that weighs 1.5 kg, is packed with information and yet makes fascinating reading from cover to cover. Mark Denny's achievement is not so much that he describes new physics or new biology but that he relates the ability of living organisms to exist, move and function to the bulk physical properties of the two substrates peculiar to Earth: air and water. Each chapter begins with a discourse on basic physics, containing just enough mathematics to enable the reader to follow the derivations of the equations that are later used to illustrate the biological consequences.

After a brief introduction, dealing with elementary concepts of dimensions, units, motion, energy and power, the majority of the book is devoted to the biophysics of fluid dynamics, diffusion, thermal, optical and electrical properties, surface tension and wave motion. The biological examples are beautifully chosen and the author displays a fine sense of humour. In the section on drag, the terminal velocity is calculated, where weight equals drag. The author comments that the terminal velocity is so called not because of the effect it has on the sky-diver (shown in an illustration) when he hits the ground, but because it is the highest velocity he can reach in free fall.

The scarcity of plankton in the air compared to water is also related to their terminal velocities and to the chaotic motion of the fluid, known as turbulence. Limits to the speed of walking also depend on terminal velocity. The author has several gentle digs at the way in which physicists manage to simplify complex problems: in a chapter on the physiological effects of wind speed, he refers to the typical 70-kg spherical human.

Inevitably there are oversimplifications and errors: water does not normally freeze at 0 °C, and atmospheric temperatures of the ecosphere are not limited to 0–40 °C. Indeed, some species depend for their survival on their ability to promote supercooling. The statement that few terrestrial animals are active below 0 °C may be correct, but activity is less important than survival, and many species can survive periods of intense cold or drought by remaining inactive. The book would have gained from a discussion of the effects of physiological water stress and adaptive

mechanisms in plants and simple forms of life; the author is mainly concerned with higher forms of life that are less able to adapt to harsh conditions.

The book is carefully crafted, well illustrated and largely devoid of typographical errors. The author even apologizes for using the letter *k* for the Boltzmann constant as well as a general-purpose coefficient; such consideration is rare. SI units are used throughout. Here is an original, amusing, well-produced and attractively priced book, ideal as a present to anyone with an ounce of natural curiosity about the fascinating interplay of physics and life on Earth. □

Felix Franks is in the Biopreservation Division, Pafra Ltd, 150 Science Park, Cambridge CB4 4GG, UK.

Random samples

Peter T. Landsberg

Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics. By Lawrence Sklar. Cambridge University Press: 1993. Pp. 437. £40, \$64.95.

THE tragedy of science is that we never know enough. That complete knowledge tends to escape us is nowhere more clearly illustrated than in statistical mechanics. As one cannot readily set up a system that follows the history of, say, 10^{23} particles, one has to eschew the search for completeness and substitute statistics. Statistical assumptions have to be made, probabilities enter and so on. As far as physics is concerned, it all started with Clerk Maxwell's distribution law in 1860. The formalism of the new science of statistical mechanics grew and grew; it came to embrace solids, liquids and gases and now the subject has several physics journals devoted to it. Even the nuclear physics community makes space for papers on it.

What are these random elements? A good way to start discussing them is to design (on paper) a coin-tossing machine, such that if a coin is placed on it in a standard way, and a button is pressed, the coin is projected upwards and always comes down heads (say). If the design is bad, or the machine ages, it will become less reliable. Lack of knowledge and lack of control has then to be replaced by probabilities. Lawrence Sklar does not give this example, but he uses the first third of the book to tell readers about probabilities and the surprisingly many ways of approaching the relevant concepts. All these various avenues are in fact left open; the author is certainly even-handed, but one would have liked some kind of summary for the sake of clarity.

In dealing with equilibrium and non-

equilibrium systems (in the second third of the book), more advanced topics make their appearance, often without, I fear, adequate explanation: KAM theorem, BBGKY approach, Bernoulli systems, non-unitary transformations. Sklar is certainly a considerable scholar, but in this book perhaps not the ideal teacher. Most welcome, however, is his introduction of the spin-echo experiments of the 1950s. Here, it will be recalled, a certain order among atomic spins is destroyed by a radio-frequency pulse. The 'coarse-grained' appearance of the system is accordingly one of equilibrium and of higher entropy relative to the initial ordered state. But lo! Another pulse restores the status quo (that is, order). This shows rather well that there was lack of equilibrium in the system at all times and that a 'fine-grained' approach is essential. In such an approach the entropy remains constant throughout. Here again the author is perhaps not as emphatic as he might be.

There is also a revealing computer experiment (not mentioned in the book) in which particles interact, the entropy is calculated, and is found to increase with time. At a certain instant all velocities are reversed and the entropy decreases in the resulting spontaneous process. This follows Loschmidt's request to Boltzmann in 1876 to see what happens to his H theorem in this case. Would he not find an entropy decrease? "You reverse them," was Boltzmann's answer (I believe), knowing that one could not reverse molecular velocities. As I say: now we can! We can also (on a computer) make a gas go into a corner of its container in an entropy-decreasing spontaneous process, by reversing the velocities that occur in the corresponding expansion. But the initial conditions for these processes are so incredibly delicate that the slightest computer errors (representing perhaps also a residual interaction with the outside world) tend to suppress this behaviour, and the system tends to go back to an entropy increase. It is the delicacy of the required initial conditions that explains (at least roughly) why an entropy decrease in an isolated system is never observed.

This takes us to the last third of the book, which deals with cosmology, the reduction of thermodynamics to statistical mechanics and time's arrow. Here a certain amount of speculation is inevitable. The author seems to favour the view that entropy continually increases in a re-contraction universe. (This depends on having at least two phases present.) There is a brief reference to arguments for the existence of God and for the author's view that entropy is not subjective, as well as some general remarks about the extension of concepts. The latter is illustrated by reference to the extension of the temperature concept to include negative temperatures.

This could have been made clearer by explicit reference to a system possessing only a finite energy spectrum.

The book occupies itself with foundations and touches on most of the crucial issues. It is, I believe, the only available modern text that has set itself this task, and as such it is recommended. While it cannot be expected to solve previously unsolved problems, it does not even classify them or bring them into sharp focus in a summary. This might be worth considering for a future edition.

The foundation problems are fairly profound; many remain unanswered, as the author points out in his conclusion, and a reader may come away with the idea that nothing very reliable can be erected on such sandy soil. This would of course be a grave error. The flow of electricity in solids, to give just one example, is often based on Boltzmann's transport equation, and is a going concern: the transport coefficients can be calculated and they often agree with experiment. While the exciting problems of the foundations of statistical mechanics is one of its glories, its widespread ability to reach out successfully to experiment is surely another which should be noted in a book of this kind. □

Peter T. Landsberg is in the Department of Mathematics, University of Southampton, Southampton SO9 5NH, UK.

New Journals issue

This year, *Nature's* annual new journals review supplement will appear in the issue of 29 September. Publishers and learned societies are invited to submit journals for review, taking note of the following criteria:

- Journals that first appeared during or after June 1992 and issued at least four separate numbers by the end of April 1994 will be considered.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine, engineering and pure mathematics are excluded, as are publications of abstracts.
- Frequency of publication must be at least three times a year. The main language used must be English. Translation journals in English are, of course, eligible.
- Deadline for submission is the end of May.

When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title, together with full details of subscription rates. For further information please contact Peter Tallack, *Nature*, 4 Little Essex Street, London WC2R 3LF, 071-836-6633 (011-44-71-836-6633 from the US), ext. 2414.

The origin and early diversification of tetrapods

Per E. Ahlberg & Andrew R. Milner

A series of new fossil discoveries, coupled with cladistic analysis of old and new data, is beginning to resolve the origin of tetrapods into a documented sequence of character acquisition. Devonian tetrapods were more fish-like than believed previously, whereas Lower Carboniferous tetrapod faunas contain early representatives of the amphibian and amniote lineages. These very different assemblages are separated by a 20-million-year 'Tournaisian gap' which has yielded very few tetrapod fossils.

THE late Palaeozoic invasion of the land by vertebrates was a key episode in the history of life, and has had a profound impact on the development of all terrestrial ecosystems. However, this event is still poorly understood. Although it has been known for a century that the terrestrial vertebrates (tetrapods) evolved from the lobe-finned fishes (sarcopterygians) during the Devonian period, only very recently has there been more direct evidence of this transition^{1,2}.

In the late nineteenth and early twentieth centuries, the earliest known fossil tetrapods were from Lower Carboniferous strata in Scotland, and were assumed to have originated during the Carboniferous, in the lowland permanent freshwater swamps that gave rise to these deposits. Barrell's theory³, of a shift onto land triggered by the periodic aridity exemplified by the Devonian red-beds, was not supported by 'hard' palaeontological data. In 1932, the tetrapod record was pushed back in the Upper Devonian by the description of the skull of *Ichthyostega*⁴ from East Greenland, and this discovery was interpreted as verification of Barrell's theory. The 363-million-year-old, metre-long *Ichthyostega* rapidly acquired symbolic status as 'the earliest land vertebrate', but it possessed a bizarre mosaic of characters, combined obvious relict fish-like structures, such as fin-rays, preopercular bones and a notochord entering the braincase, with unique features of skull and ankle construction which seemed to bar it from ancestry to later tetrapods⁵⁻⁷. This has resulted in a tendency to put *Ichthyostega* to one side and instead to look at tetrapod origins by comparing Devonian sarcopterygians with Carboniferous and Permian tetrapods.

For much of the middle of this century, there was general agreement that most of the Tetrapoda had arisen from osteolepiform lobefins (Fig. 2a-c). The principal debate centred on whether the entire Tetrapoda was monophyletic, or whether the Urodela (salamanders) had a separate origin from the others. Jarvik^{5,8,11} held to the latter view, arguing that urodeles were derived from porolepiform lobefins. Other palaeoichthyologists and palaeoherpetologists argued for monophyly¹²⁻¹⁶ and by 1981, most workers believed that all tetrapods were derived monophyletically from within or near the osteolepiforms (Fig. 1). However, while osteolepiforms are anatomically tetrapod-like in several respects, they are morphologically rather undistinguished fishes, with no unambiguous structural adaptations to an amphibious mode of life. There was thus little hard evidence to suggest how, or in what sequence, most of the key tetrapod characters had arisen. In accordance with the 'process-based' evolutionary thinking of the day, many scenarios were constructed to 'explain' the environmental background and selection pressures behind the move onto land. Most incorporated the implicit assumption that tetrapod characteristics had evolved to facilitate terrestrial life. One class of scenarios centred on the concept of vertebrates surviving drought by moving from one water body to another^{17,18}, another that vertebrates emerged

from the water either to disperse^{19,20} or to exploit the developing terrestrial ecosystem^{21,22}.

Although cladistic methodology had already been applied to problems of the interrelationships of sarcopterygian fishes^{16,23} and to tetrapod monophyly²⁴, a conceptual turning-point in the study of tetrapod origins occurred with the publication in 1981 of a large-scale cladistic review of sarcopterygian and tetrapod interrelationships²⁵, which used a wide-ranging examination of character distributions and polarities as the basis for a direct attack on the existing phylogenetic schemes and the reasoning underpinning them. Rosen *et al.*'s morphological arguments centred on the distribution of the choanae (internal nostrils) among sarcopterygian fish and lower tetrapods, but their main point was methodological. They stated categorically that evolutionary scenarios, such as those that purported to reconstruct the evolution of tetrapods from osteolepiforms, were unfalsifiable 'stories' and not testable science. They concluded that the lungfishes (Dipnoi) are the closest relatives (living and fossil) of the tetrapods, and dismissed the osteolepiforms altogether as a paraphyletic assemblage of basal sarcopterygians defined on primitive characters. Although many have strongly criticized this bold article²⁶⁻³², and its systematic conclusions have not been widely accepted, some of Rosen *et al.*'s methodological criticisms were generally perceived as justified and have consequently shifted the basis on which the topic is discussed. Nearly all subsequent papers on sarcopterygian systematics and tetrapod origins have been framed within a cladistic methodology^{16,29-33}, and most are computer-based. This has focused attention on the sequence of acquisition of tetrapod characters, which provides a more robust framework for reconstructing the base of the tetrapod tree. It also permits the reconstruction of palaeobiological models that are more constrained by character-distributions and are hence partly testable.

Despite extensive debate, no single consensus about sarcopterygian interrelationships has emerged since 1981, but some pre-cladistic theories have proved to be relatively robust. Virtually all later authors have rejected the lungfish-tetrapod proposed by Rosen *et al.*, whereas both the monophyly of the osteolepiforms and the relationship of this group to the tetrapods have been reasserted^{16,29,32-34}. Increasingly, however, the debate about tetrapod origins has focused on a previously obscure group of sarcopterygians, the family Panderichthyidae^{15,35-37}.

Panderichthyids in the Lower Frasnian

The panderichthyids, comprising the genera *Panderichthys* (Fig. 2d-f) and *Elpistostege*, are a group of early Frasnian lobe-finned fishes that in many respects resemble osteolepiforms, and indeed were originally included within that group. However, new material from Latvia and Canada has shown the Panderichthyidae to be substantially more tetrapod-like than previously realized. With some dissent^{29,30}, an increasing number of workers perceive

them as the closest known relatives of tetrapods^{15,35–37}. Indeed, one panderichthyid fragment, the holotype skull roof of *Elpistostege*, was initially described as a tetrapod³⁸, while two other supposed panderichthyids have recently proved to be Devonian tetrapods^{37,39}. The fundamental importance of panderichthyids lies in the combination of characters they possess. Unlike osteolepiforms, panderichthyids actually look like early tetrapods with paired fins: they have the same superficially crocodile-like skulls with dorsally placed orbits, straight tails and slightly flattened bodies without dorsal or anal fins (Figs 1 and 2d). Like tetrapods, but unlike all other fishes, they also have frontal bones in the skull roof (Fig. 2f)³⁵. In evolutionary terms, this simply means that a tetrapod-like body morphology appeared before the limbs. Functionally, however, it may indicate that panderichthyids had adopted a shallow-water predatory niche. The dorsally placed orbits or panderichthyids are particularly noteworthy, as this configuration is associated with aerial vision in some shallow-water teleosts, such as *Anableps* the four-eyed fish and *Periophthalmus* the mud-skipper, as well as in crocodiles. It has also recently been suggested that *Panderichthys* was capable of terrestrial locomotion similar to that in the extant catfish *Clarias*⁴⁰.

In addition to their tetrapod-like features, panderichthyids possess some unique characters (for example, the vertebral construction (Fig. 2e)) which indicate that they form a small clade rather than a paraphyletic segment of the tetrapod stem-lineage. Nevertheless, they appear to provide a Lower Frasnian 'window' on the earliest recognizable stages of tetrapod evolution. The immediately following fossil record is, however, highly 'patchy' and much of our understanding of the diversification of tetrapods is based on three subsequent 'windows' at widely spaced intervals (Fig. 3).

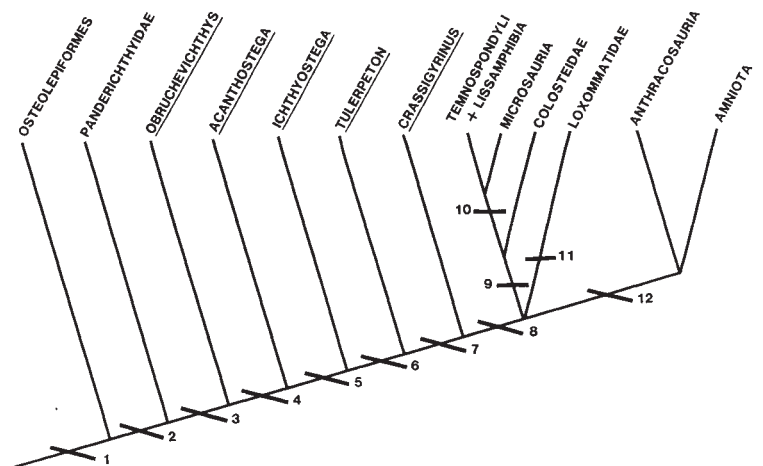
Basal stem-tetrapods

The next body of evidence consists of a range of fragments and some footprints apparently derived from true tetrapods. These date from the Upper Frasnian and Lower Famennian (middle Upper Devonian), 5–10 million years later than the panderichthyids, and have been found in Latvia (*Obruchevichthys*)^{15,37}, Scotland³⁷ and Australia (*Metaxygnathus*)^{37,41}. Most of the osteological evidence is either recently collected or recently recognized, and little can yet be said about these animals. Many of the specimens are jaw fragments, but the Scat Craig locality in Scotland has produced the earliest limb bones, from both fore and hind limbs³⁷. A humerus from this locality resembles those of tetrapods in having a thin flat entepicondyle continuous with the flat body of the humerus, and a narrow tall ectepicondyle, but in other respects it is like that of *Eusthenopteron* (Fig. 4f). A tibia from Scat Craig bears articulation surfaces for the intermedium and tibiale, suggesting the presence of a tarsus (Fig. 4h). One of the contemporaneous Australian finds is a pair of tetrapod trackways with clear footprints⁴², confirming that tetrapods were present at this time.

Primitive stem-tetrapods

After a further interval comprising most of the Famennian, a more substantial sample of early tetrapod material, including complete skeletons, appears in late Famennian horizons, representing the latest Devonian (Figs. 4a–e, g and 6). Much of this is new material, collected during the past decade. The *Ichthyostega*-bearing sediments of East Greenland have recently yielded many fossils of the previously poorly known genus *Acanthostega* which is proving to be a very different form (Fig. 4d,e)^{43–49}. A third genus, *Tulerpeton*, apparently more advanced in having relatively gracile limbs, has been collected recently in European

FIG. 1 A cladogram spanning the fish–tetrapod transition. The taxa have been grouped on the basis of the shared derived characters that they possess; the principal characters supporting each node are listed below. Inferences can be drawn about the nature of the transition from the sequence in which the derived characters (which can also be considered as evolutionary novelties) appear. The characters shared by panderichthyids and tetrapods appear to be morphological adaptations to a shallow-water habitat; Devonian tetrapods had acquired limbs and a large pelvis, but the known forms must have been semi-aquatic as they retained prominent lateral-line systems and gills; only post-Devonian tetrapods possess indisputably terrestrial adaptations, such as a strengthened vertebral column, and occipital condyles instead of a notochordal support for the skull. The move onto land could thus be interpreted as a very gradual process, with the earliest tetrapods using their limbs to move in shallow water rather than on land. This cladogram was produced iteratively rather than through computer analysis, and excludes the less completely known Devonian forms and several minor Carboniferous groups of enigmatic relationship. As the *Obruchevichthys*-grade tetrapods are poorly known, some of the characters listed at node 4 may eventually prove to be present at node 3. Significant characters defining the numbered nodes are as follows: 1, Choana present; single pair of external nares. 2, Flattened head with relatively elongate preorbital region; dorsally positioned orbits with narrow, transversely concave, interorbital region; frontal bones present; intracranial joint immobilized; body dorsoventrally flattened; absence of separate median fins; enlarged ribs. 3, Meckelian bone not exposed dorsal to the prearticular; humerus with thin, flat entepicondyle continuous with humerus body, and narrow tall ectepicondyle; tibia with articulation surfaces for intermedium and tibiale. 4, Single pair of nasals meeting in midline; compact otico-occipital region to skull; cheek with broad jugal–quadratojugal contact; absence of coronoid fangs; absence of operculars, median gular and submandibulars; presence of pre- and postzygapophyses; well developed, ventrally directed ribs; large ornamented interclavicle; carpus, tarsus and dactyly of up to eight digits; iliac blade of pelvis extending dorsally and attached to vertebral column by sacral rib; ischia contribute to pubo-



ischial symphysis; femur with adductor crest. 5, Presence of olecranon process on ulna. 6, Open lateral-line system on most or all dermal bones; elongate scapula blade with distinct cleithrum; dactyly of six digits or fewer. 7, Absence of anocleithrum. 8, Occipital condyles present; loss of preopercular; loss of anterior tectal; notochord excluded from braincase in adults; loss of ectepicondylar foramen in humerus; loss of 'd' canal in humerus; palatal dentition reduced to paired fangs (reversed in some later groups). 9, Exoccipital–postparietal contact on occiput; loss of cranial kinesis; small/incipient interpterygoid vacuities; dactyly of manus reduced to four digits or fewer. 10, Long cultriform process of the parasphenoid contacting vomers; ceratobranchial ossicles compact with only five or six denticles; narrow waisted humerus, not 'L'-shaped. 11, Antorbital vacuities at least as large as orbits. 12, Premaxillaries less than half of skull width, vomers taper anteriorly, maxillaries reach level of anterior ends of vomers; tabular–parietal contact; oblique ridge of femur reduced to ventral ridge system; pes with phalangeal formula of 2.3.4.5.4–5.

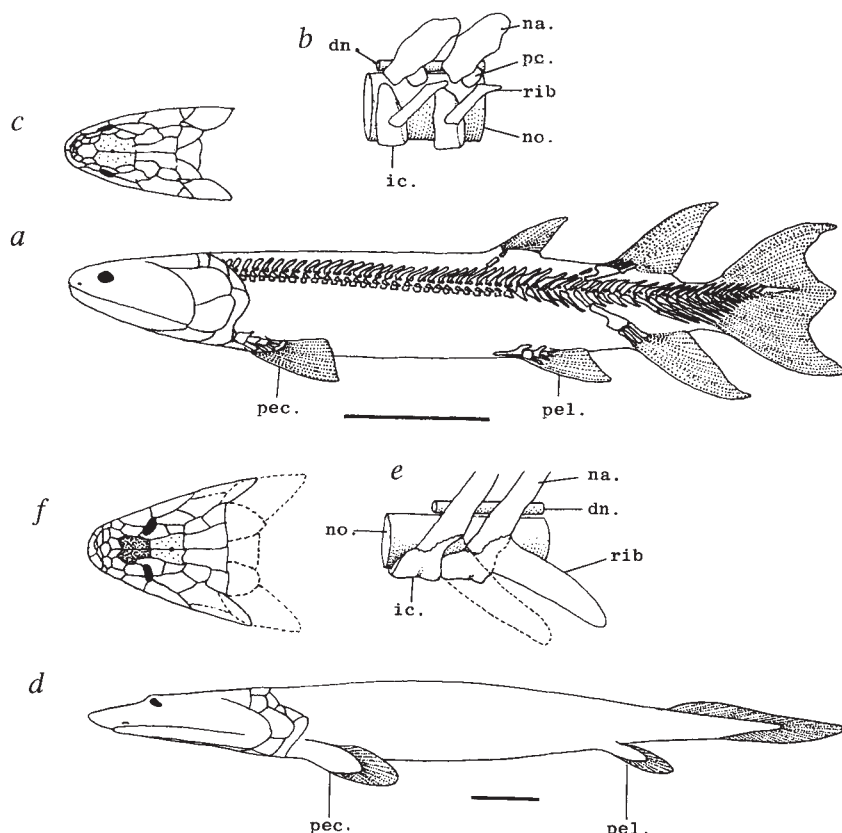
Russia⁵⁰⁻⁵², while more fragmentary tetrapod fossils from the late Famennian of Latvia are under description³⁹. *Ichthyostega*, *Acanthostega* and *Tulerpeton* seem to be morphologically disjunct representatives of a diverse basal tetrapod fauna. The latter genera relieve *Ichthyostega* of its isolated position at the base of the tetrapod tree, and give us a clearer and broader perspective of the morphological range of the most primitive tetrapods. Although they have not yet been fully described, we know that they were carnivores about 50 to 120 cm long. They differ considerably from one another, but all three genera are strikingly archaic compared with all later tetrapods except for the genus *Crassigyrinus* (see below).

The most striking feature shared by these late Famennian forms is polydactyly: *Acanthostega* had eight digits on fore (Fig. 4e) and hind limbs, *Ichthyostega* had seven on the hind limb (Fig. 4g) (forelimb unknown) (refs 47, 48 and M. I. Coates, personal communication), while *Tulerpeton* appears to have had six digits on both⁵⁰. As Coates and Clack point out⁴⁸, the discovery of these configurations is of profound importance in constraining theories of the origin and development of limbs (Fig. 5). Virtually all previous investigations have assumed that the primitive tetrapod limb pattern was pentadactyl (five-toed) and incorporated a specific arrangement of wrist and ankle bones⁵³. After comparison of this 'canonical' pattern with the skeletons of the paired fins of sarcopterygians, most workers concluded that the limb skeleton has the same basic structure as the internal skeleton of the corresponding osteolepiform paired fin (Fig. 5b)^{11,54}. Wrists, ankles and jointed digits were argued to have arisen by fragmentation and slight realignment of the distal elements of the fin skeleton. However, it now appears that neither the pentadactyl condition nor the canonical wrist pattern is primitive for tetrapods, which invalidates such earlier theories substantially if not completely. Furthermore, recently described embryological evidence⁵⁵ indicates that the digits cannot be equated with any known part of a sarcopterygian fin skeleton.

The pectoral or pelvic fin skeleton of a lobe-finned fish consists of a jointed axis carrying preaxial, and sometimes postaxial, elements (see Fig. 5a). In the living Australian lungfish *Neoceratodus*, the axial and preaxial elements arise from a single branching process during development, whereas postaxial elements develop separately as condensations of skeletal tissue (Fig. 5c,d)⁵⁵. In embryonic tetrapod limbs⁵⁵, the humerus-radius-ulna complex and the femur-tibia-fibula complex arise by a very similar branching process, the radius and tibia being preaxial branches (Fig. 5e,f). The proximal elements of tetrapod limbs thus correspond closely to those of sarcopterygian paired fins. However, the process then 'switches sides' to produce the digits as a series of postaxial branches (Fig. 5e-g). No process resembling this is known to occur or to have occurred in lobe-finned fishes. It appears that hands and feet are 'new' structures, produced by a major developmental change affecting the distal parts of the paired appendages^{47,48}.

A puzzling aspect of limb evolution is that sarcopterygian pectoral fins tend to be larger than the pelvic fins, whereas tetrapods are 'rear-wheel drive' animals with larger hind limbs than fore limbs. The earliest known tetrapods have not yet clarified the nature of this transition. In *Ichthyostega*, the fore limb is a strong prop with a permanently bent elbow (Fig. 4a)¹¹, but the fore limb of *Acanthostega* is rather feeble and much more fish-like than the hind limb (Fig. 4e)^{47,48}. The Frasnian limb bones from Scat Craig, Scotland, include a tibia (Fig. 4h) resembling those of *Ichthyostega* and *Acanthostega*, and a humerus (Fig. 4f) roughly intermediate between the osteolepiform and early tetrapod patterns³⁷. If these bones belong to the same genus—which is at present unprovable but seems likely—this suggests a degree of limb disparity even greater than in *Acanthostega*. This may indicate that the predominance of the hind limb is a very ancient tetrapod feature, and possibly that the hind limb was first to appear, although more evidence is needed before any firm conclusions can be drawn. Interestingly, the *hox* genes

FIG. 2 Osteolepiforms and panderichthyids. a-c, The Lower Frasnian osteolepiform *Eusthenopteron*. a, Lateral view showing axial skeleton and fin supports (modified from ref. 11). Scale bar, 100 mm. *Eusthenopteron* is a fusiform predatory fish with no obvious terrestrial adaptations; the paired appendages are pectoral (pec.) and pelvic (pel.) fins. As in tetrapods, there is only one pair of external nostrils. b, Section of vertebral column (modified from ref. 54); the notochord (no.) and dorsal nerve cord (dn.) have been reconstructed. The intercentra (ic.) and pleurocentra (pc.) are comparable to those of early tetrapods, but there are no articulations between the neural arches (na.), and the short ribs were probably directed dorsally. Scale bar, 10 mm. c, Skull roof; the region anterior to the parietals (light shading) is occupied by a mosaic of small bones, as in many other sarcopterygians (from ref. 11). Scale bar, 10 mm. d-f, The Lower Frasnian panderichthyid *Panderichthys rhombolepis* d, Lateral view (modified from ref. 36). Scale bar, 100 mm. *Panderichthys* is dorso-ventrally flattened, lacks separate dorsal and anal fins, and had dorsally placed eyes; these are derived characters shared with tetrapods. The axial skeleton and girdles are not completely known. e, Section of vertebral column (modified from ref. 36); there are no ossified pleurocentra, but the ribs are larger than in osteolepiforms. Scale bar, 10 mm. f, Skull roof (modified from ref. 15); panderichthyids and tetrapods both have paired frontal bones (dark shading) anterior to the parietals.



expressed in the tetrapod fore limb are interpreted by some as copies of those expressed in the hind limb^{56,57}, but others dispute this interpretation^{58,59}.

Although the Famennian tetrapods have dactylous limbs and more heavily constructed vertebral columns than lobe-finned fishes, they still appear to have been largely aquatic. Both *Ichthyostega* and *Acanthostega* retain true tail fins with fin rays. *Acanthostega* also has a large and well-ossified branchial skeleton similar to that of modern lungfishes⁴⁹, and a large supportive stapes that does not seem to have been associated with a tympanum^{45,46}. Until recently, it was widely assumed that the origin of tetrapods was synonymous with the invasion of the land, a view epitomized by countless popular illustrations of lobe-finned fishes hauling themselves laboriously onto Devonian pond margins and river banks. However, the new discoveries have led Coates and Clack^{48,49} to suggest that tetrapods evolved initially to exploit a shallow-water environment. The Famennian forms bear a broad morphological resemblance to the panderichthyids, despite the presence of limbs and longer tails, and may have filled similar niches, evolving limbs and digits to facilitate movement in weed-choked shallows rather than on land. The Sargassum frogfish (*Histrio*) provides a living example of 'underwater walking'⁴⁸. The revised view that *Acanthostega* and *Ichthyostega* were in themselves aquatic will probably not prove controversial. Whether they were primarily aquatic (that is, their ancestors had never left the water) or secondarily so (descended from more terrestrial precursors) is a more contentious issue. If *Acanthostega* was primarily aquatic, then not only did the dactylous tetrapod limb evolve underwater, but also the large tetrapod pelvic girdle and sacral rib, together with the reduction of opercular and gular ossifications.

By the end of Devonian, tetrapods had been in existence for at least seven million years and had begun to diversify. Their fossils have been found in fluvial or deltaic 'Old Red Sandstone' sediments and in lagoonal limestones, in association with fish assemblages dominated by antiarchs, osteolepiforms and lungfishes. At the Devonian–Carboniferous boundary, these assemblages cease to appear in the record, although this may represent a phenomenon of taphonomic overlay rather than sudden extinction. The vertebrate record for the ensuing 30 million years through the Tournaisian to the middle Viséan is extremely poor (Fig. 3). A few isolated bones have been collected from the Tournaisian of Nova Scotia⁶⁰, but have yet to be described. When, following this interval, the Middle Viséan–Namurian 'window' is reached, the tetrapods have diversified substantially and, except for one relict, have moved substantially beyond the grade of structural organization seen in the Famennian.

Late Lower Carboniferous tetrapod diversity

The middle Viséan–lower Namurian tetrapod fauna is known from Scotland⁶¹, Nova Scotia⁶⁰, West Virginia^{62–65}, Iowa⁶⁶, Illinois⁶⁷ and Utah⁶⁸. Most assemblages represent communities in lowland swamps or lakes on coastal plains, and all are situated in the continent of Euramerica, close to the palaeoequator of the time⁶⁹. Forms representing 15 families (effectively distinct lineages) have been or are being described and the presence of several other lineages may be deduced by phylogenetic inference. With the exception of the enigmatic *Crassigyrinus*, all the described forms are clearly of a more derived grade of organization than the Famennian tetrapods. The anthracosaurs *Eoherpeton* and *Proterogyrinus* are widely accepted to be early relatives of the amniotes^{63,70,71}, whereas the colosteids *Pholidogaster* and *Greererpeton* are probably aberrant early relatives of the modern amphibians^{65,72} (see Fig. 1 for a hypothesis of relationships of these forms). The remaining forms, the loxommatids *Loxomma*⁷³ and *Spathicephalus*, the aistopod *Lethiscus*⁷⁴, the adelogyrids⁷⁵ and the enigmatic *Caerorhachis*⁷⁶, are still of uncertain position.

For most of the last century, the Viséan–Namurian tetrapod fauna was known largely from the specialized aquatic forms

described above. It was clear that much of the Lower Carboniferous fauna, including the more terrestrial elements, was missing through taphonomic bias. The discovery in the 1980s of three new tetrapod assemblages is filling this gap. The richest of these assemblages is that discovered at East Kirkton Quarry near Bathgate in Midlothian, Scotland^{77–86}. The tetrapods were found in a local sequence of freshwater limestones, black shales and tuffs, and are associated principally with scorpions, the eurypterid *Hibbertopterus*, millepedes, a harvestman spider and a wide range of plant material. The only elements of the assemblage that seem to have been completely aquatic are abundant ostracodes. The tetrapods themselves seem to have been terrestrial forms and several of the most readily characterized forms can be assigned to groups already known from the Upper Carboniferous. These include temnospondyls, an aistopod and three members of the anthracosaur–amniote radiation^{80–84}. The common temnospondyl appears to be more advanced in skull construction than several Upper Carboniferous members of the group, and this position within the Temnospondyli implies that several other more primitive, lineages of temnospondyl were already present in the Viséan⁸⁰. The aistopod shows that this group of snake-convergent limbless tetrapods had already differentiated in the Viséan⁸¹. An amphibious anthracosaur⁸², a terrestrial anthracosaur-like form of uncertain position⁸³ and the small, long-bodied stem-amniote *Westlothiana*^{84–86}, show that the anthracosaur–amniote radiation was well differentiated on land at this time. As well as these forms, a further and as yet unpublished new tetrapod from East Kirkton combines the primitive characters of loxommatids, temnospondyls and anthracosaurs and may be close to the amphibian–amniote branching point.

The second recently discovered assemblage is that from the Viséan St Louis Formation of Delta, Iowa^{66,87}. The commonest tetrapod here is a completely new form having many of the features of anthracosaurs, but retaining a primitive tetrapod skull roof construction. It has been preliminary described as a 'protoanthracosaur', but may represent a lower grade or organ-

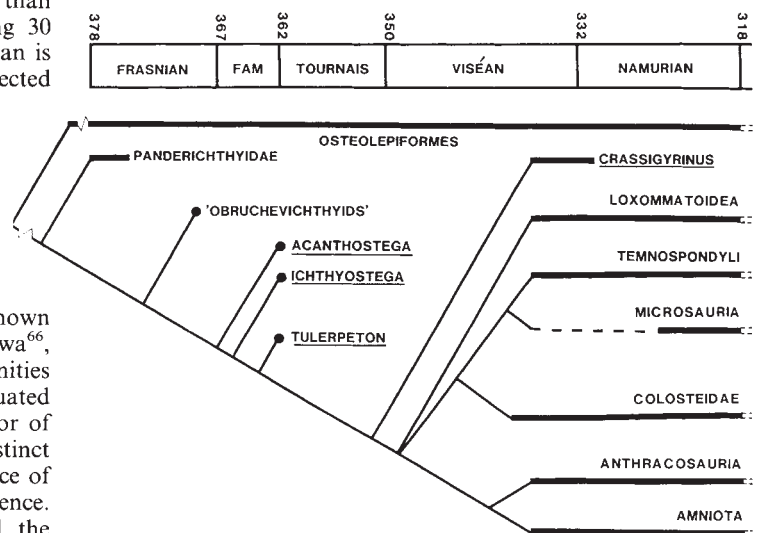
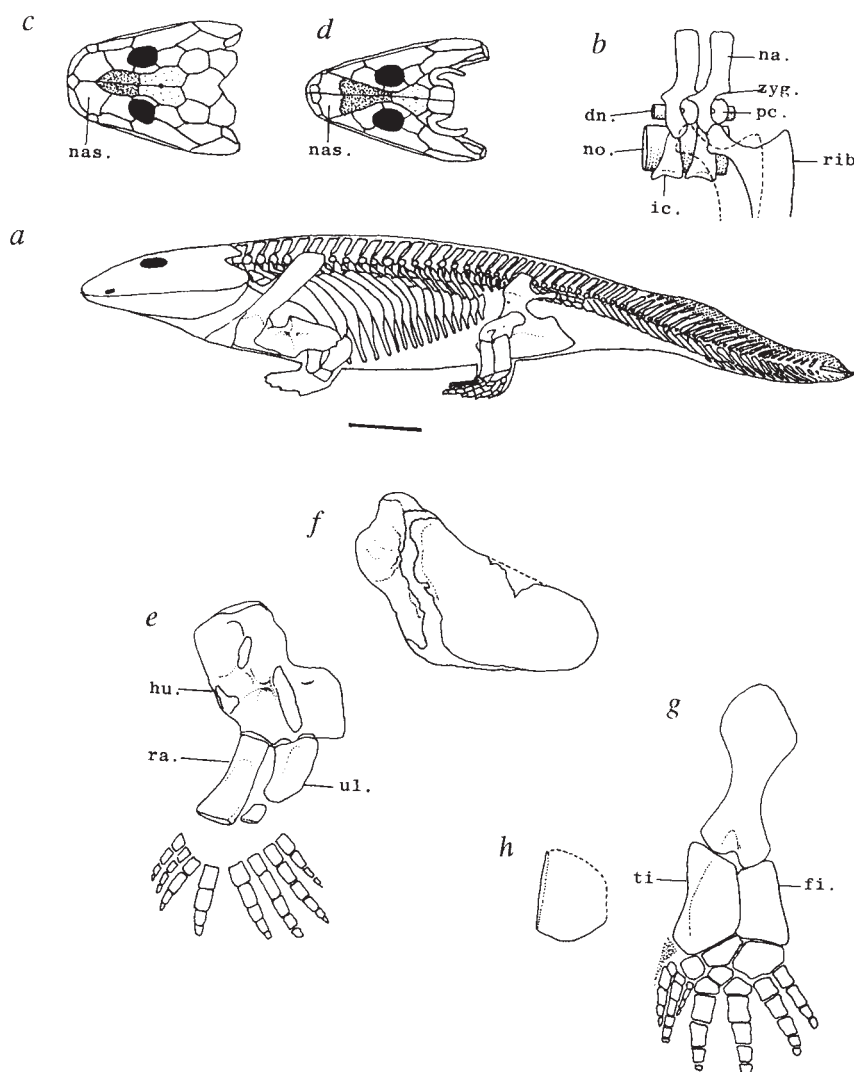


FIG. 3 The cladogram from Fig. 1 superimposed on a timescale of the Upper Devonian and Lower Carboniferous. The numbers on the left are recent datings. The black dots represent taxa known from a single locality and hence point in time; thicker bars represent the known stratigraphical range of taxa known from several records. The osteolepiforms originated before the beginning of the Middle Devonian; hence the zig-zag lines in the lowermost part of the cladogram indicate a time span of at least 15 million years. Seven of the groups extend on beyond the Namurian. The diagram emphasizes the large chronological gaps between the known assemblages through this period of time. FAM, Famennian; TOURNAIS, Tournaisian.

FIG. 4 Devonian tetrapods. *a*, Reconstruction of the Upper Famennian genus *Ichthyostega* in lateral view (modified from ref. 11). This is the only whole-body reconstruction yet published for a Devonian tetrapod; note the large pelvic girdle, the rib-cage composed of broad overlapping ribs, and the loss of the dorsal series of bones connecting the pectoral girdle with the head. The fore limb has a permanently flexed elbow, but the structure of the manus is unknown. The tail carries bony fin rays similar to those in *Panderichthys*. Scale bar, 100 mm. *b*, Section of vertebral column (modified from ref. 11); the general structure is similar to that of *Eusthenopteron*, but the neural arches (na.) interlock by means of zygapophyses (zyg.), and the ribs are much larger. Scale bar, 10 mm. *c,d*, Skulls of *Ichthyostega* (from ref. 11) and *Acanthostega* (from refs 43, 44) in dorsal view; all known Devonian tetrapods possessed frontals (dark shading) and paired nasals (nas.), *e*, Fore limb of the Upper Famennian tetrapod *Acanthostega* in dorsal view (from ref. 48); unlike *Ichthyostega*, this animal had a straight elbow with limited mobility. There are eight digits, and the proportions of the radius (ra.) and ulna (ul.) are very fish-like, although the humerus (hu.) resembles those of other early tetrapods. Coates and Clack⁴⁸ regard this as the most primitive known tetrapod fore limb. *f*, Upper Frasnian humerus from Scat Craig, Scotland (from ref. 37); this humerus, about seven million years older than *Acanthostega*, has more fish-like proportions. *g*, Hind limb of *Ichthyostega* in dorsal view (after ref. 48); the foot has seven digits preceded by a poorly ossified mass and the tibia (ti.) and fibula (fi.) are strikingly broad. *h*, Upper Frasnian tibia from Scat Craig (from ref. 37); this incomplete tibia is closely comparable to that of *Ichthyostega*, but seems to be proportionately shorter. Scale bars for *e,f* and *g,h*, 10 mm.



ization. It is represented by several large complete skeletons, somewhat crushed, and should provide the head-to-tail osteology of a tetrapod close the amphibian-amniote dichotomy. It is accompanied by as colosteoid amphibian and fishes.

Third, the basal Namurian Manning Canyon Shale near Utah Lake in Utah has produced two specimens of a microsauro, together with some of the earliest known winged insects⁶⁸. The microsauro, *Utaherpeton*, can be attributed to the Microbrachiomorpha, one of the subgroups recognized in later microsaurs. Thus, like the East Kirkton material, it takes the diversification of an Upper Carboniferous group well back in the Lower Carboniferous.

The systematics of early tetrapods

For many years, the basal tetrapods were conveniently placed in two large groups, the Labyrinthodontia characterized by labyrinthine teeth and multipartite vertebral centra, and the Lepospondyli characterized by simple teeth and unipartite centra. The advent of cladistic methodology has left the Labyrinthodontia exposed as a paraphyletic group based solely on primitive characters^{70,88}, while the Lepospondyli is more controversially seen as a polyphyletic assemblage of dwarf early tetrapods^{1,70,88,89}. Consequently, the use of these terms is declining but no consensus has yet emerged on the groupings that may replace Labyrinthodontia and Lepospondyli. For many workers the problem can be seen in terms of the living Tetrapoda. There is widespread (though not uniform) agreement that the Lissamphibia (the living amphibians) and the Amniota (the reptiles, birds and mammals) are each monophyletic groups. If this is accepted, it follows that early tetrapods could be representatives

of, or sidelines off, the stem of the Lissamphibia, the stem of the Amniota, or the common tetrapod stem. To place early tetrapods into one of these three categories would be considerable progress, not only for the systematist, but also for the palaeobiologist wishing to understand the sequence of character acquisition in early tetrapods.

Figure 1 represents a set of relationships for the tetrapod taxa discussed above and below. No consensus exists, but this proposal is based largely on published arguments^{65,78,88}. *Acanthostega* and *Ichthyostega* are perceived as the most adequately known primitive stem-tetrapods, with *Crassigyrinus* (discussed below) further up the stem. The loxomatids are ambiguous in their position and could be either stem-tetrapods or stem-amniotes. Other early tetrapods fall within the amphibian/lissamphibian clade (the microsaurs, colosteids and temnospondyls), or the amniote clade (the anthracosaurs and true amniotes). The discovery of all these groups in the Viséan-Namurian sequences shows that the radiation of the 'crown-group' tetrapods was well under way.

Crassigyrinus. *Crassigyrinus*, from the Viséan and Namurian of Scotland, is a bizarre and systematically challenging form which has recently generated much discussion^{90,91}. It was a basically eel-like form with tiny limbs but was at least 130 cm long (Fig. 7). The short-snouted skull had deep massive cheeks, prominent labyrinthodont teeth and fangs; *Crassigyrinus* was clearly a large aquatic predator. There are two conflicting interpretations of its systematic position. Panchen^{90,92} has argued that,

although an extremely primitive tetrapod, *Crassigyrinus* shares significant characteristics with the anthracosaurs, which in turn are generally accepted as primitive amphibious relatives of the amniotes (reptiles, birds and mammals). In this view, *Crassigyrinus* would be the most primitive representative of the line leading to amniotes and would show that the division between the two lineages giving rise to the modern amphibians and to the amniotes had occurred at a very basal, possibly pre-terrestrial stage in tetrapod evolution. The alternative view^{36,65,78,93} is that *Crassigyrinus* lacks so many characteristics shared by other Carboniferous and later tetrapods that it must be an early offshoot of the tetrapod stem, close to *Acanthostega* and *Ichthyostega* but convergently resembling the anthracosaurs.

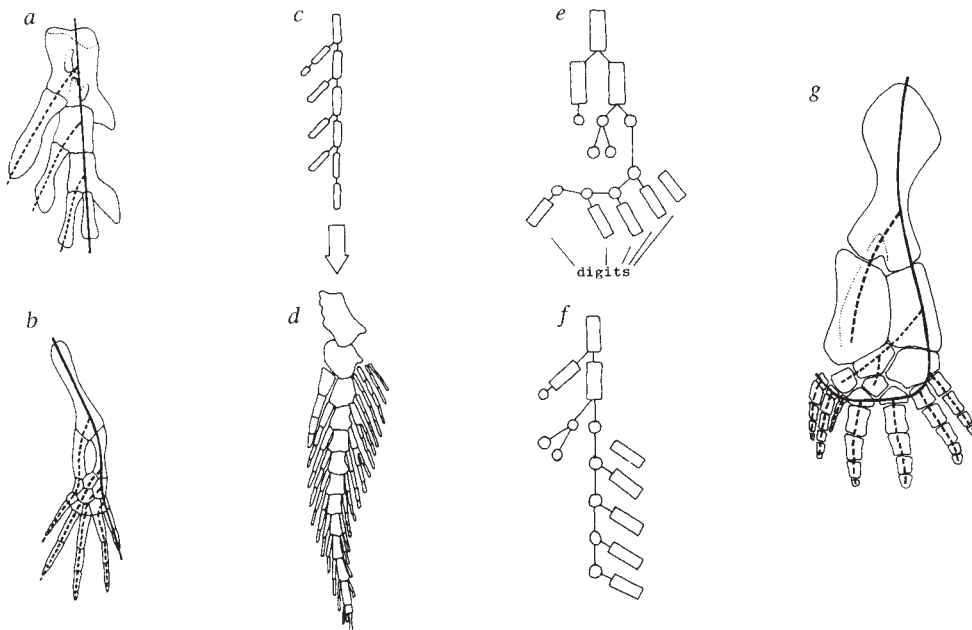
The new interpretations of *Acanthostega* and *Ichthyostega*, as primary aquatic forms that had not yet become terrestrial, raises the possibility that *Crassigyrinus* is not only a structural relict but also an ecological relict. Instead of being the first tetrapod to 'return to the water', it may be the last survivor of the primitive tetrapods that never left the water.

Basal crown-group tetrapods. The loxommatids, a group of early tetrapods known since the last century, are uniquely characterized by curious keyhole-shaped orbits⁷³. The loxommatids clearly form a compact group, but their extrinsic relationships are enigmatic. They were previously considered to be temnospondyls and hence a side-branch on the lineage leading to modern amphibians. Recently it has been shown that their similarity to temnospondyls is based entirely on primitive tetrapod features, and some derived similarities to anthracosaurs—and hence amniotes—have been proposed^{70,88,94}. Others have suggested that loxommatids belong on the tetrapod stem above *Crassigyrinus*⁷⁸. The limited information available from loxommatid skulls seems insufficient to place them precisely, but they appear to be very close to the amphibian–amniote dichotomy. A new loxommatid specimen from the Upper Carboniferous of Lancashire, UK, has the first partial postcranial skeleton associated with this family and should generate considerable data about their position⁹⁵.

Another form that may belong to this grade of organization is *Caerorhachis*. Described as a very primitive temnospondyl with anthracosaur-like vertebrae, this incomplete specimen could equally be a juvenile loxommatid, a basal temnospondyl or a basal anthracosaur⁸⁰. Two of the most recent discoveries that also fall in this category are the Delta 'protoanthracosaur' and the Kirkton archaic tetrapod. Neither is a loxommatid, but both are almost universally primitive in the characteristics they share with temnospondyls, loxommatids and anthracosaurs. This cluster of forms, when fully described, will give us a basic morphotype for the tetrapod crown-group. Whereas the Devonian tetrapods and *Crassigyrinus* clearly had only a partial set of tetrapod characteristics, these forms appear to have most, if not all of them, and are close to the point at which the modern amphibians and amniotes diverged.

Stem-amphibians. The true amphibians (that is, all those fossil forms that are more closely related to living amphibians than to living amniotes) are widely perceived as a group that were successful in the Carboniferous but later were outcompeted by the amniotes for all but a restricted range of terrestrial niches. This is a somewhat exaggerated view, not only because of the immense diversity of modern amphibians (more than 4,000 species), but also because the Carboniferous record largely derives from ponds, abandoned channels and ox-bow lakes, in which amphibians might be expected to be better represented than amniotes. Nevertheless, early members of the amphibian clade, such as the large aquatic colosteids, the small terrestrial, burrowing and aquatic microsaur and the crocodile- and salamander-like temnospondyls, were certainly diverse in the Upper Carboniferous. Recent discoveries are tending to push this diversity down into the Lower Carboniferous. The Viséan colosteid *Pholidogaster* was described in the last century, but only recognized as a colosteid in 1975⁹⁶. The first Namurian microsaure from the Manning Shale of Utah was described in 1990⁶⁸. This form, *Utaherpeton*, can be assigned to one of the subgroups of microsaur, the Microbrachomorpha, and suggests some diversification by that time. An undescribed assemblage from

FIG. 5 Paired fins and limbs. a, Pectoral fin skeleton of the osteolepiform *Eusthenopteron*, a relatively short structure consisting of an axis (solid line) carrying preaxial radials (dashed lines). Phylogenetic evidence suggests that the tetrapod limb is derived from this general type of fin skeleton. b, Gegenbaur's⁹⁶ plan of an idealized limb, showing inferred position of the axis and radials. In this interpretation (and those of most subsequent authors), the tetrapod digits correspond to preaxial radials, and the overall structure closely resembles the fin skeleton of *Eusthenopteron* (a and b after ref. 48). c, d, Pectoral fin skeleton of the Recent lungfish *Neoceratodus*. c, Diagrammatic representation of early development (after ref. 55). The axis and preaxial radials are laid down by a branching process; the fin skeleton of *Eusthenopteron* was probably formed in the same way. d, Adult fin. The numerous postaxial radials form as condensations of tissue, not through branching. e, Fore limb skeleton of mouse, *Mus musculus*; diagrammatic representation of early development (after ref. 55). The proximal part of the limb skeleton consists of an axis with preaxial branches (radials), as in *Eusthenopteron* (a) or embryonic *Neoceratodus* (c), but the digits are formed as postaxial branches. This pattern is unknown in fishes. f, Limb skeleton of *Mus musculus* straightened out to facilitate comparison with fin skeletons. g, Hind limb of the Devonian tetrapod *Ichthyostega*, showing inferred position of axis (solid line) and radials (dashed lines), following ref. 55 (after ref. 48). Other known Devonian tetrapod limbs have eight (*Acanthostega*) or six (*Tulerpeton*) digits. All limbs/fins shown with anterior edge to left.



letons. g, Hind limb of the Devonian tetrapod *Ichthyostega*, showing inferred position of axis (solid line) and radials (dashed lines), following ref. 55 (after ref. 48). Other known Devonian tetrapod limbs have eight (*Acanthostega*) or six (*Tulerpeton*) digits. All limbs/fins shown with anterior edge to left.

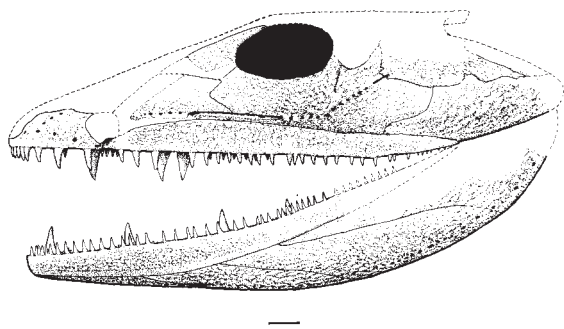


FIG. 6 Skull reconstruction of *Ventastega curonica*, a recently discovered Devonian (Upper Famennian) tetrapod from Latvia (after ref. 39). Scale bar, 10 mm.

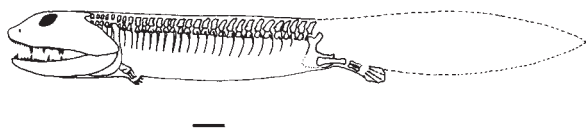


FIG. 7 The primitive Carboniferous (Viséan and Namurian) tetrapod *Crassigyrinus*. This large predator had tiny fore limbs and must have been wholly aquatic (modified from refs 90 and 91). Scale bar, 100 mm.

the Viséan of Illinois also includes what appear to be microsauroid vertebrae⁶⁷. The common temnospondyl at East Kirkton is an unexpectedly advanced form and implies that several temnospondyl families previously known only from later horizons must also have had antecedents in the Viséan⁸⁰.

Stem-amniotes. The anthracosaurs are widely accepted to be stem-amniotes, and were known to exist in the Viséan and Namurian, based on such forms as *Eoherpeton* from Scotland and *Proterogyrinus* from Scotland and West Virginia, USA. The East Kirkton fauna appears to include a variety of stem-amniotes, one of which can be readily characterized as a true anthracosaur⁸², but other forms are also present. The enigmatic *Westlothiana*, initially reported as the earliest reptile^{84–86}, may prove to be slightly further down the amniote stem, but still appears to be a post-anthracosaurian offshoot. Yet another Kirkton tetrapod also seems to be a terrestrial stem-amniote, but of less certain position⁸³. Again, the principal significance of the new material is, at this stage, that it shows that much of the diversity previously associated with Upper Carboniferous faunas was already present in the Viséan.

A Tournaisian radiation?

As new information from these four sets of assemblages, from panderichthyids to late Lower Carboniferous tetrapod faunas, emerges, it appears increasingly that there was an immense radiation of terrestrial tetrapods in the Tournaisian. On the one hand, Viséan and early Namurian faunas are proving to be more advanced than previously realized, with temnospondyls, anthracosaurs, aistopods, microsaurs and possible amniotes, and thus resembling later Carboniferous faunas in their taxonomic composition. On the other hand, the sequence of Lower Frasnian panderichthyids, Upper Frasnian *Obruchevichthys*-grade forms and Famennian tetrapods looks increasingly like a series of 'state of the art faunas', and suggests that basal tetrapod morphology was being acquired during the Upper Devonian, but that the latest Devonian forms may not yet have been fully terrestrial. Thus, more advanced tetrapod morphologies are being described from the Viséan, while Famennian tetrapods prove to be more

primitive than previously suggested. The Tournaisian and early Viséan may thus be the critical time both for the functional water-land transition and the ensuing diversification of the truly terrestrial tetrapods.

The alternative hypothesis is that tetrapod diversification did occur gradually during the Famennian, and that forms such as *Acanthostega* and *Ichthyostega* are already relicts when we see them in the late Famennian. Such a theory permits a more generous time for the diversification of the tetrapods present by the Upper Viséan, but the only direct evidence to support it is Famennian postcranial material of *Tulerpeton*⁵⁰ which is not yet fully described. This is clearly more derived than that of the other Devonian tetrapods, and resembles that of later terrestrial forms in certain respects.

Future research directions

We can expect a substantial increase in information about each of the three tetrapod assemblages reviewed above. New material of the Upper Frasnian *Obruchevichthys*-grade from Scotland will be described and further collecting at a productive locality is under way. For the Famennian tetrapods, detailed descriptions of *Acanthostega*, *Ichthyostega* and a new Latvian genus (Fig. 6) are in press or preparation, and a collaborative Anglo-Russian study of *Tulerpeton* in conjunction with the Greenland material is being undertaken, details of the cranial component having just been published⁵². Most of the nineteenth-century material from the classical Viséan-Namurian localities has been further prepared and redescribed over the past two decades, but description of articulated tetrapod skeletons in the East Kirkton, Delta and Manning Canyon Shale assemblages will add substantially to the known diversity of amphibious-terrestrial tetrapods for this interval. The most interesting forms are proving to be those nearest the base of the tetrapod crown-group, and the description of the Delta 'protoanthracosaur', the East Kirkton archaic tetrapod and the later Wigan loxomatid postcranium will together provide us with a view of tetrapod construction at, or close to, the point when the ancestors of modern amphibians and the amniotes diverged. These forms are probably relicts of the 'state-of-the-art' in the Tournaisian and the nearest we can come to understanding tetrapod evolution at that time. The next challenge will be to find such Tournaisian tetrapod assemblages. □

Per E. Ahlberg is at the Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK and Andrew R. Milner is at the Department of Biology, Birkbeck College, Malet Street, London, WC1E 7HX, UK.

1. Carroll, R. L. *Am. J. Sci.* **20**, 45–84 (1992).
2. Thomson, K. S. *Am. J. Sci.* **293A**, 33–62 (1993).
3. Barrell, J. *Bull. geol. Soc. Am.* **27**, 387–436 (1916).
4. Sæve-Søderbergh, G. *Meddr. Grønland* **94**, 1–107 (1932).
5. Jarvik, E. *Zool. Bidr. Upps.* **21**, 235–675 (1942).
6. Jarvik, E. *Ark. Zool.* **41A**, No. 13, 1–8 (1948).
7. Jarvik, E. *Meddr. Grønland* **114**, 1–90 (1952).
8. Jarvik, E. *Scient. Mon.* **80**, 141–154 (1955).
9. Jarvik, E. *Colloq. int. Cent. natn. Rech. Scient.* **104**, 87–101 (1962).
10. Jarvik, E. *Kung. Svensk. Vetensk. Handl.* (4) **9**, 1–74 (1963).
11. Jarvik, E. *Basic Structure and Evolution of Vertebrates* vol. 1 (Academic, London, 1980).
12. Parsons, T. S. & Williams, E. E. *Q. Rev. Biol.* **38**, 26–53 (1963).
13. Thomson, K. S. *Nobel Symp.* **4**, 285–305 (1968).
14. Jurgens, J. D. *Ann. Univ. Stellenbosch A* **46**, 1–146 (1971).
15. Vorobyeva, E. I. *Trudy paleont. Inst.* **163**, 1–239 (1977).
16. Schultze, H.-P. *Paläont. Z.* **55**, 71–86 (1981).
17. Romer, A. S. *Vertebrate Paleontology* 3rd edn (Chicago Univ. Press, 1966).
18. Schmalhausen, I. I. *The Origin of Terrestrial Vertebrates* (Academic, New York, 1968).
19. Ewer, D. W. *Science* **122**, 467–468 (1955).
20. Feduccia, J. A. *Texas J. Sci.* **22**, 255–263 (1971).
21. Gunter, G. *Science* **123**, 495–496 (1956).
22. Cox, C. B. *Proc. Linn. Soc. Lond.* **178**, 37–47 (1967).
23. Andrews, S. M. in *Interrelationships of Fishes* (eds Greenwood, P. H., Miles, R. S. & Patterson, C.) 137–177 (Academic, London, 1973).
24. Gaffney, E. S. *Bull. Carnegie Mus. nat. Hist.* **3**, 92–105 (1979).
25. Rosen, D. E., Forey, P. L., Gardiner, B. G. & Patterson, C. *Bull. Am. Mus. nat. Hist.* **167**, 159–276 (1981).
26. Jarvik, E. *Syst. Zool.* **30**, 378–384 (1981).
27. Holmes, E. B. *Biol. J. Linn. Soc.* **25**, 379–397 (1985).
28. Schultze, H.-P. in *The Biology and Evolution of Lungfishes* (eds Bemis, W. E., Burggren, W. W. & Kemp, N. E.) 39–74 (Liss, New York, 1987).

29. Panchen, A. L. & Smithson, T. R. *Biol. Rev.* **62**, 341–438 (1987).
30. Chang, M.-M. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 3–28 (Cornell, Ithaca, NY, 1991).
31. Ahlberg, P. E. *Zool. J. Linn. Soc.* **96**, 119–166 (1989).
32. Ahlberg, P. E. *Zool. J. Linn. Soc.* **103**, 241–287 (1991).
33. Schultze, H.-P. *Verh. dt. zool. Ges.* **84**, 135–151 (1991).
34. Long, J. A. *J. Vert. Paleont.* **9**, 1–17 (1989).
35. Schultze, H.-P. & Arsenault, M. *Palaeontology* **28**, 293–309 (1985).
36. Vorobyeva, E. & Schultze, H.-P. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 68–109 (Cornell, Ithaca, NY, 1991).
37. Ahlberg, P. E. *Nature* **354**, 298–301 (1991).
38. Westoll, T. S. *Nature* **141**, 127 (1938).
39. Ahlberg, P. E., Lukševičs, E. & Lebedev, O. *Phil. Trans. R. Soc. B.* **343**, 303–328 (1994).
40. Vorobyeva, E. & Kuznetsov, A. in *Fossil Fishes as Living Animals* (ed. Mark-Kurik, E.) 131–140 (Acad. Sci. Estonia, Tallinn, 1992).
41. Campbell, K. S. W. & Bell, M. W. *Alcheringa* **1**, 369–381 (1977).
42. Warren, J. W. & Wakefield, N. A. *Nature* **238**, 469–470 (1972).
43. Clack, J. A. *Palaeontology* **31**, 699–724 (1988).
44. Clack, J. A. *Geol. Today* **4**, 192–194 (1988).
45. Clack, J. A. *Nature* **342**, 425–430 (1989).
46. Clack, J. A. in *The Evolutionary Biology of Hearing* (eds Webster, D. B., Fay, R. R. & Popper, A. N.) 405–420 (Springer, New York, 1992).
47. Coates, M. I. in *Developmental Patterning of the Vertebrate Limb* (eds Hinchliffe, J. R., Hurler, J. M. & Summerbell, D.) 325–337 (Plenum, New York, 1991).
48. Coates, M. I. & Clack, J. A. *Nature* **347**, 66–69 (1990).
49. Coates, M. I. & Clack, J. A. *Nature* **352**, 234–235 (1991).
50. Lebedev, O. A. *Dokl. Akad. Nauk SSSR* **278**, 1470–1473 (1984).
51. Lebedev, O. A. *Priroda* **1985**, 26–36 (1985).
52. Lebedev, O. A. & Clack, J. A. *Palaeontology* **36**, 721–734 (1993).
53. Watson, D. M. S. *Anat. Anz.* **44**, 24–27 (1913).
54. Andrews, S. M. & Westoll, T. S. *Trans. R. Soc. Edinb.* **68**, 207–329 (1970).
55. Shubin, N. H. & Alberch, P. *Evol. Biol.* **20**, 319–387 (1986).
56. Tabin, C. J. *Development* **118**, 289–296 (1992).
57. Tabin, C. & Laufer, E. *Nature* **361**, 692–693 (1993).
58. Coates, M. *Nature* **364**, 195–196 (1993).
59. Thorogood, P. & Ferretti, P. *Nature* **364**, 196 (1993).
60. Carroll, R. L., Belt, D., Dineley, D. L., Baird, D. & McGregor, D. C. 24th int. Geol. Congr., Montreal, Field Excursion A59: Guidebook 1–113 (McAra, Calgary, Canada, 1972).
61. Smithson, T. R. *Scott. J. Geol.* **21**, 123–142 (1985).
62. Smithson, T. R. *Zool. J. Linn. Soc.* **76**, 29–90 (1982).
63. Holmes, R. *Phil. Trans. R. Soc. B* **306**, 431–527 (1984).
64. Godfrey, S. J. *Kirtlandia* **43**, 27–36 (1988).
65. Godfrey, S. J. *Phil. Trans. R. Soc. B* **323**, 75–133 (1989).
66. Bolt, J. R. *Nat. Geogr. Res.* **6**, 339–354 (1990).
67. Schultze, H.-P. & Bolt, J. R. *Spec. Pap. Palaeontology* (in the press).
68. Carroll, R. L., Bybee, P. & Tidwell, W. D. *J. Paleont.* **65**, 314–322 (1991).
69. Milner, A. R. in *Palaeozoic Vertebrate Biostratigraphy and Biogeography* (ed. Long, J. A.) 324–353 (Belhaven, London, 1993).
70. Smithson, T. R. *Zool. J. Linn. Soc.* **85**, 317–410 (1985).
71. Smithson, T. R. *Palaeontology* **29**, 603–628 (1986).
72. Milner, A. R. *Herpetol. Monogr.* **7**, 8–27 (1993).
73. Beaumont, E. H. *Phil. Trans. R. Soc. B* **820**, 29–101 (1977).
74. Wellstead, C. F. *Palaeontology* **25**, 193–208 (1982).
75. Andrews, S. M. & Carroll, R. L. *Trans. R. Soc. Edinb., Earth Sci.* **82**, 239–275 (1991).
76. Holmes, R. & Carroll, R. L. *Bull. Mus. comp. Zool. Harv.* **147**, 489–511 (1977).
77. Wood, S. P., Panchen, A. L. & Smithson, T. R. *Nature* **314**, 355–356 (1985).
78. Milner, A. R., Smithson, T. R., Milner, A. C., Coates, M. I. & Rolfe, W. D. I. *Mod. Geol.* **10**, 1–28 (1986).
79. Rolfe, W. D. I. et al. *Geol. Soc. Am. Spec. Pap.* **244**, 13–24 (1990).
80. Milner, A. R. & Sequeira, S. E. K. *Trans. R. Soc. Edinb., Earth Sci.* **84**, 331–361 (1994).
81. Milner, A. C. *Trans. R. Soc. Edinb., Earth Sci.* **84**, 363–368 (1994).
82. Clack, J. A. *Trans. R. Soc. Edinb., Earth Sci.* **84**, 369–376 (1994).
83. Smithson, T. R. *Trans. R. Soc. Edinb., Earth Sci.* **84**, 377–382 (1994).
84. Smithson, T. R., Carroll, R. L., Panchen, A. L. & Andrews, S. M. *Trans. R. Soc. Edinb., Earth Sci.* **84**, 383–412 (1994).
85. Smithson, T. R. *Nature* **342**, 676–678 (1989).
86. Smithson, T. R. & Rolfe, W. D. I. *Scott. J. Geol.* **26**, 137–138 (1990).
87. Bolt, J. R., McKay, R. M., Witzke, B. J. & McAdams, M. P. *Nature* **333**, 768–770 (1988).
88. Panchen, A. L. & Smithson, T. R. in *The Phylogeny and Classification of the Tetrapods Vol. 1: Amphibians, Reptiles, Birds* (ed. Benton, M. J.) 1–32 (Clarendon, Oxford 1988).
89. Thomson, K. S. & Bossy, K. H. *Forma functio* **3**, 7–31 (1970).
90. Panchen, A. L. *Phil. Trans. R. Soc. B* **309**, 505–568 (1985).
91. Panchen, A. L. & Smithson, T. R. *Trans. R. Soc. Edinb., Earth Sci.* **81**, 31–44 (1990).
92. Panchen, A. L. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 110–144 (Cornell, Ithaca, NY, 1991).
93. Clack, J. A. in *Spec. Pap. Palaeontology* (in the press).
94. Panchen, A. L. in *The Terrestrial Environment and the Origin of Land Vertebrates* (ed. Panchen, A. L.) 319–350 (Academic, London, 1980).
95. Milner, A. C. & Lindsay, W. *First World Congress of Herpetology, Abstr.* 201 (1989).
96. Panchen, A. L. *Phil. Trans. R. Soc. B* **259**, 581–640 (1975).
97. Gegenbaur, C. *Morph. Jb.* **2**, 396–420 (1876).

ARTICLES

Cerium/lead and lead isotope ratios in arc magmas and the enrichment of lead in the continents

Daniel M. Miller^{*†}, Steven L. Goldstein[†] & Charles H. Langmuir^{*}

^{*} Lamont-Doherty Earth Observatory and Department of Geological Sciences, Columbia University, Palisades, New York 10964, USA

[†] Max-Planck-Institut für Chemie, Postfach 3060, D-55020 Mainz, Germany

Lead is anomalously enriched in the Earth's continental crust, and in the magmas that give rise to new continental crust at convergent margins. A detailed study of volcanic rocks from the Aleutian island arc shows that this enrichment is a continuing process, and results from the efficient non-magmatic transfer of mantle-derived lead into the source of convergent-margin magmas. This process, acting through time, can also account for the pervasive depletion of lead in the oceanic mantle.

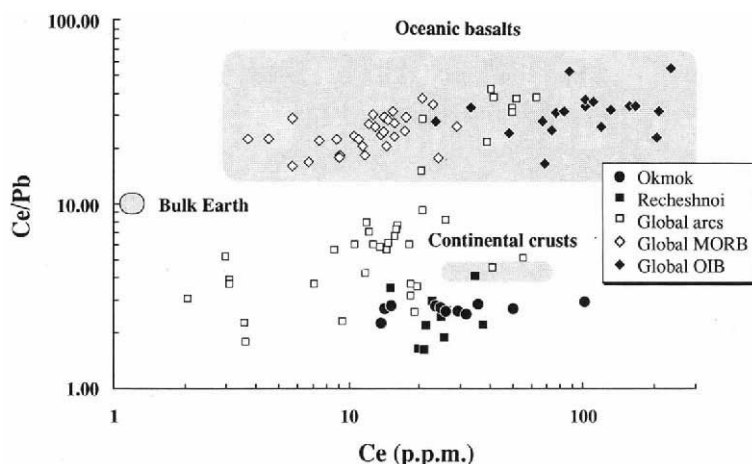
MOST elements with similar solid-melt distribution coefficients have similar concentration ratios in the bulk silicate Earth, the continental crust, and basalts formed by mantle melting at mid-ocean ridges (MORB) and oceanic islands (OIB)^{1–3}. The moderately incompatible element pair cerium and lead (Ce and Pb) is an exception. Despite their similar melting behaviours, Pb is preferentially partitioned into the continents, with enrichments similar to those of highly incompatible elements like rubidium and uranium (Rb and U)^{3–5}. Thus, compared with the bulk Earth, the Ce/Pb ratio is low in the continents and high in oceanic basalts (Fig. 1).

Suggested explanations of the preferential partitioning of Pb into the continental crust include: (1) Pb behaves magmatically

as a highly incompatible element during continent formation, but as a moderately incompatible element during generation of oceanic basalts⁵; (2) extra Pb is added to the continents by non-magmatic (for example, hydrothermal) processes⁴; (3) Pb was enriched in the continents primarily during the Archaean and early Proterozoic by processes different from the present day⁵; and (4) Pb is carried from the subducted oceanic crust to the sources of continent-building magmas by fluids before melting^{5–9}.

These explanations can be further evaluated through the study of convergent margin volcanoes, which represent recent additions to the continental crust. Most arc lavas have low Ce/Pb, similar to the continents (Fig. 1), but like oceanic basalts they

FIG. 1 Global comparison of Ce/Pb ratios in oceanic basalts, convergent margin lavas, the continental crust and the bulk Earth. Logarithmic axes are used in order to balance relative variations. Data for Okmok and Recheshnoi volcanoes (on Umnak island, Aleutian Islands) are from this work. All other data are taken from literature. Only samples with Pb concentration determined by isotope dilution or spark source mass spectrometry are included. Mid-ocean ridge basalt (MORB) and oceanic island basalt (OIB) data from refs 4, 5, 57. Arc data from refs 17, 53, 58–63. The only arc data extending into the field of oceanic basalts are from the high-calcium series lavas from Grenada⁶³.



are derived principally from melting of the mantle. New trace-element and Pb isotope data for lavas from two neighbouring Aleutian volcanoes (Table 1) place additional constraints on the origin of their low Ce/Pb ratios, and reveal a process that can explain the preferential enrichment of Pb in the continents. The volcanoes' proximity reduces variations in tectonic factors that may affect the composition of arc lavas, for example, the amount and composition of subducted sediment and the age of the subducting lithosphere, thus comparison of these sample suites presents a more controlled experiment than is possible by studying isolated lavas from widely separated volcanoes. We show that Ce and Pb behave similarly during mantle melting to form arc basalts, as they do during the formation of MORB and OIB^{3,5}, but that mantle Pb is preferentially transferred to the arc magma source before melting. The global fractionation of Pb from Ce reflects non-magmatic enrichment processes associated with the creation and subduction of oceanic crust over the Earth's history.

Geological background

Okmok and Recheshnoi, adjacent volcanoes on Umnak island, Aleutian Islands, Alaska, are separated by only 45 km along the volcanic front of the arc. They are, however, chemically distinct, following classic tholeiitic and calc-alkaline differentiation paths, respectively¹⁰. Okmok is an active shield volcano that experienced a caldera-forming eruption approximately 2,500 years ago¹¹. Our samples are lavas and pyroclastics associated with its most recent phases of volcanism¹², including historic eruptions. Recheshnoi is a smaller, glacially-eroded stratovolcano. Our samples represent both pre- and post-glacial phases of volcanism¹². Miller *et al.*¹⁰ have shown that the major- and trace-element differences between parental lavas from the two volcanoes can be accounted for by different degrees of partial melting of a similar mantle source. The samples from Okmok appear to be related by differentiation of parental magmas, whereas those from Recheshnoi require more complex relationships. The overall difference in extent of melting between Okmok and Recheshnoi provides an exceptional opportunity to investigate the behaviour of Pb in arc magmas.

Analytical results and implications

Umnak lavas exhibit linear correlations between Pb isotope ratios and Ce/Pb (Fig. 2), with most of the variation represented by Recheshnoi lavas. Okmok lavas have nearly constant Pb isotope and Ce/Pb ratios among rock types ranging from basalt to rhyolite. The uniform Pb isotope ratios in Okmok lavas indicate that the lavas are related by differentiation of melts derived from a homogeneous source, as suggested by Miller *et al.*¹⁰. In Recheshnoi lavas, Ce/Pb correlates with Pb isotope ratios (Fig.

2c), but not with concentrations (Fig. 3d), suggesting that variations in Ce/Pb result primarily from heterogeneity in the mantle source, and not from magmatic processes. The simple relationships among Okmok lavas and more complicated ones among lavas from Recheshnoi are further supported by excellent correlations of Pb abundances with Ce, U and Rb abundances in the Okmok suite ($R^2 = 0.995$, 0.970 and 0.986, respectively) and poorer correlations for the same element pairs in the Recheshnoi suite ($R^2 = 0.286$, 0.684 and 0.747).

The Okmok data reveal the behaviour of Pb during magmatic differentiation in comparison to that of Ce, U and Rb (Fig. 3a–

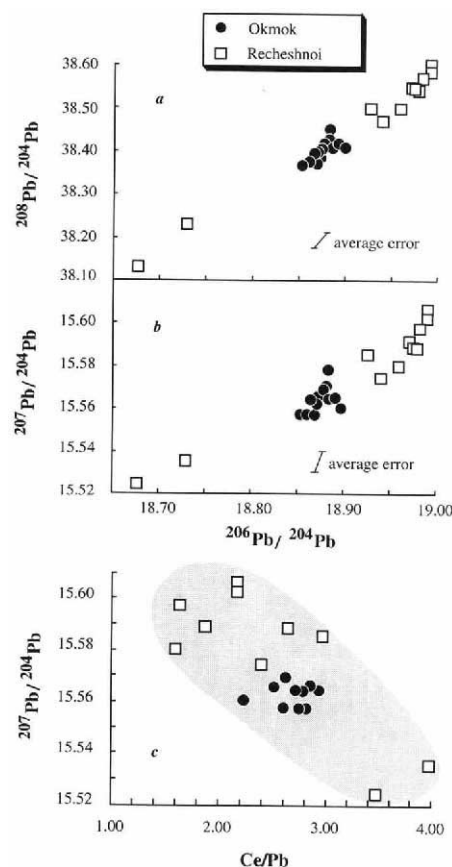


FIG. 2 Lead isotope ratios for lavas from Umnak island. a, Correlated variations in the $^{208}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ ratios. b, Variations in the $^{207}\text{Pb}/^{204}\text{Pb}$ ratio with the $^{206}\text{Pb}/^{204}\text{Pb}$ ratio. c, Correlation between $^{207}\text{Pb}/^{204}\text{Pb}$ and Ce/Pb ratios. Average 2σ uncertainty of 0.21‰ per a.m.u. is illustrated by the error bars in a and b.

TABLE 1 Analyses of lava from two Aleutian island volcanoes

Okmok volcano									
Sample	SiO ₂	MgO	Ce	Pb	U	Rb	²⁰⁶ Pb/ ²⁰⁴ Pb	²⁰⁷ Pb/ ²⁰⁴ Pb	²⁰⁸ Pb/ ²⁰⁴ Pb
87-03-01	53.39	4.10	26.04	9.97	1.23	26.63	18.869	15.557	38.370
87-06-01	55.55	3.30	30.48	—	—	35.36	18.884	15.578	38.449
87-08-02	56.99	2.90	33.17	—	1.46	44.05	18.872	15.562	38.392
87-08-09	57.52	2.78	36.09	12.57	1.79	39.90	18.874	15.566	38.401
87-17-02	70.85	0.10	102.85	34.92	3.83	102.69	18.871	15.564	38.397
Qaf70	51.35	5.13	15.32	5.42	0.61	13.81	18.861	15.557	38.372
Qcb1	52.11	4.36	31.93	12.63	1.46	31.16	18.893	15.565	38.415
Qcb7	52.55	4.09	23.55	8.40	1.01	22.92	18.867	15.564	38.393
Qcb12	54.50	3.22	50.98	18.68	2.41	51.17	18.886	15.564	38.406
Qcb16	53.19	4.08	29.43	11.22	1.34	27.89	18.878	15.569	38.418
Qcf3	50.05	5.06	14.26	5.23	0.53	11.00	—	—	—
Qdf1	50.05	6.08	13.85	6.18	0.53	10.57	18.899	15.560	38.409
Qef1	54.67	3.69	24.90	8.99	1.20	26.93	18.854	15.557	38.365
Recheshnoi volcano									
LUM16	61.12	3.16	24.80	10.34	1.75	39.81	18.941	15.574	38.468
LUM17	57.82	3.98	20.29	12.34	1.54	35.87	18.984	15.597	38.568
LUM18	54.84	4.69	21.35	—	—	35.45	18.973	15.591	38.546
LUM19*	59.25	3.82	35.35	8.88	2.27	30.92	18.730	15.535	38.227
LUM20	63.65	1.51	37.41	17.23	2.87	69.74	18.992	15.602	38.585
LUM21	50.07	10.10	15.18	4.35	0.51	12.18	18.678	15.524	38.130
LUM22	63.07	1.85	23.17	7.79	1.20	31.67	18.928	15.585	38.498
LUM33	59.16	3.82	25.46	13.58	2.00	48.46	18.977	15.589	38.540
LUM37	54.93	7.67	21.31	13.32	2.11	39.54	18.961	15.580	38.497
LUM42	60.71	3.20	21.57	9.92	1.78	43.64	18.992	15.607	38.598
ABY9(23)†	57.40	4.00	26.30	9.93	2.04	51.21	18.981	15.588	38.539

Analysis techniques: SiO₂, MgO and Ce were measured by direct current plasma emission spectrometry at Lamont-Doherty and reported in ref. 10 along with other major and trace elements. Rb, U and Pb concentrations were measured by isotope dilution at the Max-Planck-Institut, Mainz on unleached powders, and Pb isotope ratios on leached whole-rock chips for Okmok samples and leached powders for Recheshnoi samples. Leaching was in hot 6N HCl for 2 h. Chemical techniques were similar to those described previously at MPI-Mainz^{10,17,53,54}. Pb isotope ratios are corrected for mass fractionation by $\sim 1.5\%$ per atomic mass unit (a.m.u.), based on multiple analyses of the NBS 982 standard. 2σ errors (external) on the standard were 0.45‰ ($n=16$) and 0.26‰ ($n=15$) per a.m.u. during two measuring periods. Each sample was measured at least twice in order to ensure reproducibility of the analyses. 2σ errors (external) on samples were 0.21‰ per a.m.u., excluding two outliers (Qdf1 and LUM20), whose inclusion brings the error to 0.27‰ per a.m.u. Further analytical details are presented in ref. 39.

* Major- and trace-element data for sample LUM19 are not given in ref. 10, but are available from D.M.M. on request.

† SiO₂, MgO and Ce for sample ABY9(23) are from ref. 55.

c). The Ce/Pb ratio varies within a narrow range, between 2.5 and 3 in all but one sample, over nearly an eightfold range of Ce concentrations. This constancy holds also for Recheshnoi, albeit with greater scatter (Fig. 3d). In contrast, Rb/Pb and U/Pb increase with increasing concentrations in both volcanoes. These results show that Pb and Ce have similar solid-melt partition coefficients during crystallization, whereas Rb and U are more incompatible, as is true for MORB and OIB^{4,5}. The Ce/Pb ratio is therefore largely unaffected by shallow-level differentiation processes at convergent margins.

Source enrichment or melting?

The Ce/Pb ratios of all Umnak samples are within the range of island arc and continental rocks globally, and are a factor of 5–10 lower than in oceanic basalts (Fig. 1). Can the low Ce/Pb be produced by fractionation during mantle melting, or must the Umnak magma source have low Ce/Pb before melting?

We can place limits on the fractionation of Ce/Pb during melting from theory and observation. The maximum enrichment of Pb during fractional fusion would occur if Pb were perfectly incompatible, that is, with a solid-melt distribution coefficient, D_{Pb} , of zero. D_{Ce} is probably less than 0.05 (ref. 10). But even with $D_{Pb}=0$ and D_{Ce} as high as 0.1, the extent of melting required to generate melts with Ce/Pb=2.5 (typical of Umnak lavas) would be very small ($\leq 1\%$) if the magma source had Ce/Pb=25 (typical of oceanic basalts). More realistic values of D_{Ce} would imply even smaller melt fractions of $\leq 0.5\%$. Such small extents of melting are inconsistent with the observed major-element compositions of primitive arc lavas¹³, which require ~ 5 –25% melting to explain the global range^{14,15}.

The lack of Ce/Pb fractionation during melting is also illus-

trated by the Umnak data. Miller *et al.*¹⁰ have shown that Okmok and Recheshnoi can be related by different extents of melting of broadly similar source compositions, with $\sim 20\%$ melting for Okmok and $\sim 7\%$ for Recheshnoi. If Pb were more incompatible than Ce, smaller extents of melting would lead to lower Ce/Pb, but instead the Ce/Pb ratios in Recheshnoi lavas bracket those of Okmok. The U/Pb and Rb/Pb ratios in Recheshnoi lavas are higher than in Okmok (Fig. 3), consistent with U and Rb being more incompatible than Pb during melting.

This observation also holds for the global arc data set. Fractionation-corrected Ce and sodium abundances ($Ce_{6.0}$ and $Na_{6.0}$) correlate with one another, and much of the global range in these values can be explained by different extents of melting¹⁵. However, smaller extents of melting (higher $Na_{6.0}$) do not correlate with lower Ce/Pb ratios.

Therefore, from theory and observation, within a single arc and globally, it is clear that Pb is not significantly fractionated from Ce by partial melting. Cerium and Pb have similar behaviours during mantle melting in all modern tectonic environments. The low Ce/Pb in arc volcanics must result from a ubiquitous enrichment of Pb in convergent margin magma source regions. Cerium itself is enriched in the Umnak source relative to MORB mantle¹⁰, therefore, the Pb enrichment must be even greater to explain the low Ce/Pb.

Origin of the extra Pb

To understand the enrichment process, it is necessary to identify the origin of the extra Pb in Umnak lavas. From isotope data alone, the sources of Aleutian magmas have been successfully modelled as mixtures of mantle peridotite, with low ²⁰⁷Pb/²⁰⁴Pb, and small amounts (1–3%) of subducted sediment having

high $^{207}\text{Pb}/^{204}\text{Pb}$ (refs 16, 17). It has also been suggested¹⁸ that the low Ce/Pb in arc lavas may reflect such a subducted sediment component.

Our Pb isotope data provide further evidence that the low $^{207}\text{Pb}/^{204}\text{Pb}$ source component is mantle-derived. These data form a linear array between north Pacific MORB (taken to represent the Pacific oceanic mantle) and the average of sediments from Deep Sea Drilling Project (DSDP) Hole 183 (ref. 19), just outside the Aleutian trench near Umnak (Fig. 4a). On the basis of the Pb isotope ratios of the Pacific mantle and DSDP sediments ($^{206}\text{Pb}/^{204}\text{Pb} = 18.3\text{--}18.6$ and $^{207}\text{Pb}/^{204}\text{Pb} \approx 19.2$, respectively), between 40 and 60% of the Pb in Okmok lavas is derived from the 'mantle component'.

Whereas the Pb isotope data clarify the relative amounts of mantle and sediment Pb in Umnak lavas, the Ce and Pb concentration data provide new constraints on the source of the mantle Pb. Because the Ce/Pb has not been significantly fractionated by melting or crystallization, it closely reflects the magma source composition. If addition of sediment to the mantle wedge underlying the arc was entirely responsible for the low Ce/Pb, then the Umnak lavas should lie on a mixing array between Pacific mantle and sediment in a plot of $^{207}\text{Pb}/^{204}\text{Pb}$ versus Ce/Pb. Such a mixing line clearly does not pass through the Umnak data (Fig. 4b). Even samples having low $^{207}\text{Pb}/^{204}\text{Pb}$, with their Pb derived primarily from mantle sources, have Ce/Pb approximately a factor of five lower than oceanic basalts. Thus, the low Ce/Pb in Umnak lavas cannot be explained solely by addition of sediment to the mantle wedge, but requires enrichment of the magma source by a component with mantle-derived Pb and low Ce/Pb before melting.

Additional constraints on the source of this component result from the correlation between Ce/Pb and $^{207}\text{Pb}/^{204}\text{Pb}$ (Fig. 2c). If this correlation is interpreted as a mixing array, one end-member has high $^{207}\text{Pb}/^{204}\text{Pb}$ and low Ce/Pb, and probably represents subducted sediment. The second end-member, which is the dominant source of mantle Pb in Umnak lavas, must have Ce/Pb < 6 even with $^{207}\text{Pb}/^{204}\text{Pb}$ as low as 15.45, an isotope ratio that clearly reflects a mantle source. This low $^{207}\text{Pb}/^{204}\text{Pb}$, low Ce/Pb source component could result from fluid-phase extraction of Pb from the subducted MORB crust. Fluid phases may preferentially extract Pb relative to Ce from the oceanic crust as a result of (1) metamorphic dehydration of the subducted basalt^{20,21}, and (2) hydrothermal circulation through the basaltic crust at the mid-ocean ridge, which pre-concentrates MORB Pb near the surface of the oceanic crust in metalliferous sediments and sulphides before subduction^{8,9,22-25}, increasing its potential for contribution to the arc magma source. Alternatively, fluid derived from the subducted crust might preferentially extract mantle Pb if it continuously equilibrates with the mantle wedge as it migrates toward the melting region^{26,27}. If the mantle-derived component has $^{206}\text{Pb}/^{204}\text{Pb} \approx 18.47$ and $^{207}\text{Pb}/^{204}\text{Pb} \approx 15.50$ (typical of north Pacific MORB), then Ce/Pb is ~ 4 in this component (Fig. 4b) and it provides $\sim 40\%$ of the Pb in Okmok lavas. Thus the Umnak lavas provide strong evidence for selective enrichment of mantle-derived Pb, and there are reasonable mechanisms to generate the appropriate component and transfer it into the arc magma source region.

Secular depletion of Pb in the mantle

We have seen that the Umnak data place new constraints on the present-day fluxes of Ce and Pb at convergent margins. Preferential transport of mantle Pb relative to Ce into arc magmas is required. This produces arc crust with low Ce/Pb, and increases the Ce/Pb ratio in the residual mantle from which the extra Pb is derived. Therefore the Ce/Pb fractionation observed in the Aleutians today operates in the correct sense to account for the difference between the mantle and continental crust. Could subduction processes acting through geological time produce the global fractionation of Pb from Ce between the Earth's mantle and continents?

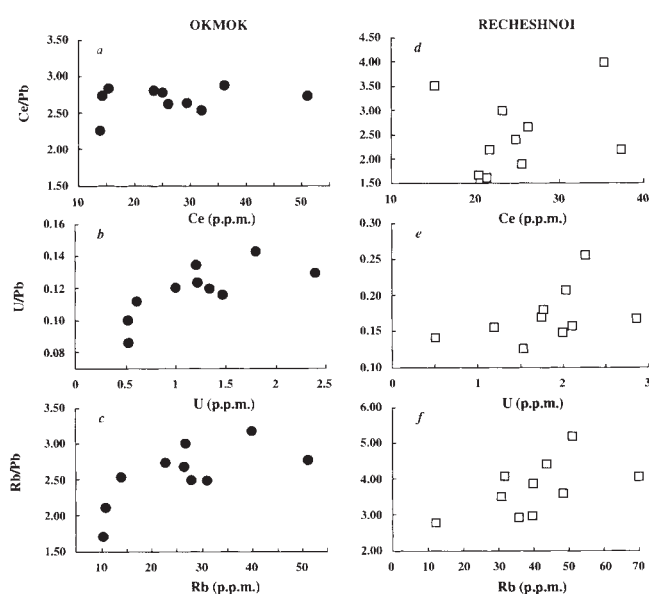


FIG. 3 Variation of the ratios of Ce, U and Rb to Pb during crystallization of the Umnak island lavas. a and d, Ce versus Ce/Pb for Okmok (excluding rhyolite sample 87-17-02) and Recheshnoi volcanoes, respectively. b and e, U versus U/Pb. c and f, Rb versus Rb/Pb. The vertical axes are scaled to show the same relative range of variation within each volcano.

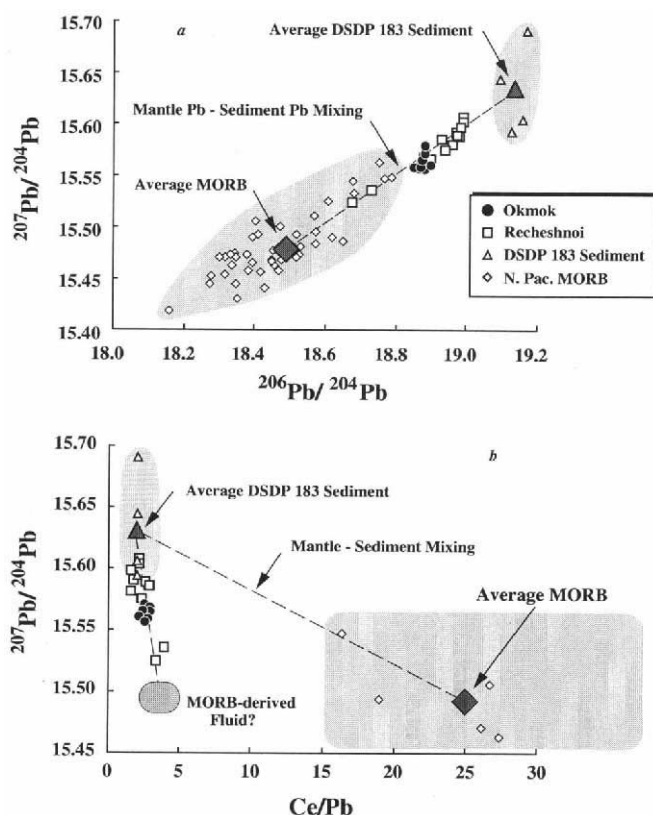


FIG. 4 Constraints on Umnak source components derived from Ce/Pb and Pb isotope ratios. a, Comparison of the Pb isotope ratios of Umnak lavas with those of north Pacific MORB and oceanic sediments just outside the Aleutian trench (Deep Sea Drilling Project Hole 183). b, Adds Ce/Pb elemental ratios to the comparison and shows that a component with low $^{207}\text{Pb}/^{204}\text{Pb}$ and low Ce/Pb is required to explain the Umnak data. Pb isotopes are shown for DSDP 183 sediment samples A2 7B4, A2 9A1 and A2 10A4 (from ref. 19) and sample 19-183-38-3, 20-21 (B. Peucker-Ehrenbrink, personal communication). Average Ce/Pb ratio for Pacific pelagic clays is from ref. 18.

We begin to address this question with a quantitative evaluation of current element fluxes from the mantle to arc magmas. The calculations assume that the mantle-derived Pb in the Okmok magma source comes entirely from the subducted oceanic crust. Although this component may be derived from both the subducted oceanic crust and overlying mantle wedge, this is unimportant to the elemental budgets over time as long as the subducted oceanic crust and wedge are mixed back into the mantle. Assuming that (1) parental arc magmas have 11.1 p.p.m. Ce (the average of parental magmas for Okmok and Rechesnoi¹⁰) and 4.42 p.p.m. Pb (thus Ce/Pb=2.5, like the average in Umnak lavas); (2) the mass of arc magma generated is 1/20 the mass of subducted oceanic crust^{28,29}; and (3) the portion of mantle-derived Pb in the magma is 40% (as in Okmok), then the concentration of Pb in the oceanic crust would be reduced by 0.088 p.p.m. (4.42 p.p.m. $\times$ 0.05 $\times$ 0.40). If Ce/Pb is $\sim$ 4 in the mantle-derived component of arc magmas (as in Umnak, Fig. 4b), the concentration of Ce in the residual MORB would be reduced by 0.354 p.p.m. (4 $\times$ 0.088 p.p.m.).

Whereas the average Ce/Pb ratio of the bulk oceanic crust is well-constrained to be $\sim$ 25 (refs 4, 5), estimates of the abundance of these elements range widely from 7.5–12 p.p.m. Ce and 0.30–0.49 p.p.m. Pb (refs 3, 30). The higher estimate³ represents an average of 26 MORB glasses, including differentiated lavas. The lower estimate assumes a primary MORB with 10% MgO (W. McDonough, personal communication), which we accept as more appropriate for the bulk oceanic crust. If oceanic crust has 7.5 p.p.m. Ce and 0.30 p.p.m. Pb when formed, the fluxes of Ce and Pb observed in the Umnak source would reduce the Ce concentration by 4.7% (0.354 $\div$ 7.5 = 0.047) and the Pb by 29% (0.088 $\div$ 0.30 = 0.29). The Ce/Pb ratio in the subducted oceanic crust would therefore increase from 25 to 34 as a consequence of preferential Pb extraction during the present subduction cycle. Over time, this process would lead to a secular increase in the Ce/Pb of the mantle.

To assess the quantitative effect on the mantle through geological time, our calculations assume that the subduction process has always produced the fractional reductions in Ce and Pb concentrations in subducted oceanic crust that are indicated today, that the 'interactive mantle' (the portion depleted to form continents) is the source of MORB and has approximately constant mass, and that the remainder of the mantle is undepleted. Using the refractory element ratio of Ce to U in the primitive mantle from ref. 3 and calculating the U/Pb ratio of the bulk Earth assuming μ (²³⁸U/²⁰⁴Pb) = 8.5 $\pm$ 0.5 and κ (²³²Th/²³⁸U) = 4.0 $\pm$ 0.2, the Ce/Pb ratio of the primitive mantle is 10.7 $\pm$ 0.5 (in general agreement with corrections to estimates in refs 3 and 30; A. Hofmann and W. McDonough, personal communication). A simple mass-balance model (Box 1) can then be used to examine the effects of processing the interactive mantle through convergent margins as oceanic crust to produce the increase of Ce/Pb from $\sim$ 10.7 to $\sim$ 25 in the present 'depleted mantle' source of MORB and most OIB^{3,4}.

The continental crust resulting from such convergent margin processing has a Ce/Pb ratio of $\sim$ 2.5 when Ce/Pb in the depleted mantle reaches 25. If we select a mass for the continental crust, equation (6) (Box 1) defines the requisite mass of the depleted mantle for any concentration of Ce in the continental crust, C_{cc}^{Ce} . Estimates of the Ce abundance in the bulk continental crust range from 33 to 55 p.p.m. (ref. 31). Using mantle and crustal masses from refs 32–34, generation of this range of Ce concentrations would require the depleted mantle to include 73–121% of the total mantle. For the mass of the depleted mantle, M_{dm} , to be $\leq$ 100% of the total mantle, Ce in the continental crust would have to be $\leq$ 45 p.p.m. Given an average age of $\sim$ 60 Myr for present-day oceanic crust³⁵, subduction of the mass of oceanic crust needed to process the interactive mantle over the age of the Earth would require average crustal production rates 2–3 times those of the present day. This requirement may be a natural consequence of the secular cooling of the Earth^{36–39}.

Although this simple model can account for the global fractionation of Pb from Ce in the silicate Earth, the results are not in complete agreement with some generally held views of the present-day Earth. The flux of Ce from the mantle into the crust is too low relative to the flux of Pb, generating continents with Ce/Pb = 2.5, lower than most estimates, which range from 3.8 to 5.0 (refs 31, 32). This in turn causes the Ce/Pb of the mantle

BOX 1 Ce/Pb mass-balance model

THE flux of an element, i , from the mantle into the oceanic crust can be described in terms of the cumulative mass of oceanic crust subducted over time, M_{oc} , and the concentration of i in the interactive mantle, C_m^i , by

$$dM_{oc}^i = EC_m^i dM_{oc} \quad (1)$$

where E is a melting function that relates C_m^i to the concentration of i in the oceanic crust ($C_{oc}^i = EC_m^i$). A fraction of this flux, f^i , is removed from the oceanic crust at convergent margins and added to the continental crust. The remaining fraction is returned to the mantle by subduction of the oceanic crust. The irreversible flux of element i out of the mantle is thus equal to the flux of i into the continental crust,

$$-M_{dm} dC_m^i = f^i EC_m^i dM_{oc} \quad (2)$$

where M_m is the mass of the interactive mantle. Formally, the mass of the interactive mantle is the sum of the masses of the continental crust, M_{cc} , and the residual depleted mantle, M_{dm} . Because the mass of the continental crust is only 0.5% of the silicate Earth, to simplify the model we use the approximation $M_m \approx M_{dm}$. Thus, $M_{dm} \approx M_{m0}$, the initial mass of the interactive mantle. Accordingly, the depleted mantle is hereafter assumed to represent the interactive mantle. Rearranging equation (2) and integrating yields:

$$C_{dm}^i \approx C_{m0}^i \exp \left(-\frac{M_{oc}}{M_{dm}} f^i E \right) \quad (3)$$

where C_{m0}^i is the concentration of i in the primitive mantle. The ratio, R , of two elements i and j in the depleted mantle is then given by

$$R_{dm}^{i/j} \approx R_{m0}^{i/j} \exp \left(\frac{M_{oc}}{M_{dm}} E (f^j - f^i) \right) \quad (4)$$

where $R_{m0}^{i/j}$ and $R_{dm}^{i/j}$ are the ratios in the primitive and depleted mantle, respectively. We assume that E represents equilibrium melting with partition coefficients for Ce and Pb of 0.01, and that mid-ocean ridge basalts (MORBs) are produced by an average of 10% melting ($F=0.1$) (ref. 56). If $f^{Ce}=0.047$ and $f^{Pb}=0.29$ as is presently observed for the volcanoes on Umnak island, then to increase the Ce/Pb ratio of the interactive mantle from its primitive value, $R_{m0}^{Ce/Pb} \approx 10.7$, to its present value, $R_{dm}^{Ce/Pb} \approx 25$, requires M_{oc}/M_{dm} to be $\sim$ 0.38. The total mass of mantle melted over geological time to form MORB is M_{oc}/F . Thus to satisfy the model, the interactive mantle would have to be processed $\sim$ 4 times (0.38 $\div$ 0.1), through creation and subduction of oceanic crust.

Substituting $M_{oc}/M_{dm}=0.38$ into equation (3), with $C_{m0}^{Ce}=1.60$ (ref. 3) and $C_{m0}^{Pb}=0.150$, the concentrations of Ce and Pb in the present-day depleted mantle would be 1.36 p.p.m. and 0.055 p.p.m., respectively. Because these calculations assume that the depleted mantle is the complement of the continental crust, the mass of element i in the continents, M_{cc}^i , is given by

$$M_{cc}^i = M_{m0}^i - M_{dm}^i \quad (5)$$

where M_{m0}^i is the initial mass of i in the interactive mantle and M_{dm}^i is the mass of i in the depleted mantle. Because one of our approximations is that $M_{dm} \approx M_{m0}$, equation (5) can be written as

$$\frac{M_{cc}}{M_{dm}} \approx \frac{(C_{m0}^i - C_{dm}^i)}{C_{cc}^i} \quad (6)$$

where M_{cc} is the mass of the continental crust and C_{cc}^i is the concentration of element i in the continental crust. Substituting the equivalent mass-balance approximation for j into equation (6), we obtain

$$\frac{C_{cc}^i}{C_{cc}^j} \approx \frac{C_{m0}^i - C_{dm}^i}{C_{m0}^j - C_{dm}^j} \quad (7)$$

to increase to 25 before a sufficient mass of Ce and Pb has been extracted, yielding concentrations of Ce (1.36 p.p.m.) and Pb (0.055 p.p.m.) in the depleted mantle that are too high relative to present-day MORB sources. With the high mantle abundances generated by the model, mass balance between the crust and mantle (equation (7), Box 1) also requires a depleted mantle that is larger than most previous estimates^{5,31,36,37}. Thus, the simple model generates continents with Ce/Pb that is too low, and a depleted mantle that is too large and has Ce and Pb abundances that are too high.

More realistic models of mantle and crustal evolution would have more free parameters, accounting for recycling of continental crust, alternative mechanisms of crustal growth and mantle depletion, and exchange between the MORB mantle and a less-depleted mantle reservoir. Intramantle exchange^{31,38}, for example, would relax the mass-balance constraints on the size and composition of MORB mantle by distributing the depletion throughout the entire mantle. However, the general success of our simple model illustrates the importance of convergent margin processes for global Ce/Pb fractionation. Although rigorous evaluation of the effects of all simplifying assumptions is beyond the scope of this Article, we will briefly consider some of their implications.

Recycling of continental sediments into the mantle would reduce the rate of depletion caused by continent formation, increasing the total mass of oceanic crust that would have to be processed to satisfy the model. But Nd and Pb concentration and isotope data for Umnak lavas^{39,40} and DSDP Hole 183 sediments^{19,41,42} show that Pb is transferred preferentially relative to Nd from subducted sediments into the arc magma source. Okmok lavas have Nd/Pb ≈ 1.7 , with $\sim 60\%$ of the Pb and $\sim 10\%$ of the Nd derived from subducted sediment. Thus the sediment-derived component returned to the continents through arc magmatism has Nd/Pb ≈ 0.3 ($1.7 \times 0.1 \div 0.6$), a factor of ~ 5 lower than in DSDP-183 sediments. Residual sediment that is recycled to the deep mantle would therefore have higher Nd/Pb (and by inference, higher Ce/Pb) than bulk subducted sediment. Such sediment recycling would inhibit mantle depletion (defined as the decrease in Ce or Pb abundance), but for a given degree of depletion it would lead to higher mantle Ce/Pb and lower crustal Ce/Pb. Depending on the mass of sediment subducted relative to the mass of crust produced, a proportion that could vary greatly over Earth history, sediment recycling could enhance, retard or even reverse the Ce/Pb fractionation described in the simple model.

The flux of Ce to the continents may be increased relative to the flux of Pb by other continent-building mechanisms, such as intraplate magmatism and basaltic underplating^{29,43–46}. Because these mechanisms reflect mantle melting without involving the processes that fractionate Pb from Ce, the continental crust that they generate would have high Ce/Pb, similar to the mantle source. The effect of such mechanisms can be modelled by increasing the extraction efficiencies f^{Ce} and f^{Pb} (See Box 1) by equal amounts. Increasing each f by 14.5% yields a reasonable Ce concentration of 0.82 p.p.m. in the depleted mantle, but implies a rather high Ce/Pb of ~ 6.7 for the continental crust. Thus, intraplate magmatism or underplating might increase the flux of Ce from the mantle sufficiently to yield a reasonable MORB source, but would generate continents with Ce/Pb higher than published estimates.

It is also possible that higher mantle temperatures in the past may have resulted in more widespread melting of the subducted oceanic crust. Because melts would transfer Ce and Pb from the oceanic crust into the arc source with equal efficiency, f^{Ce} and f^{Pb} would be increased by equal amounts. Therefore, the effect on mantle depletion would be similar to that of intraplate magmatism.

An additional process that would deplete Ce and Pb in the MORB mantle, without significantly affecting the Ce/Pb of the continental crust, is the accumulation of subducted oceanic litho-

sphere in a boundary layer within the mantle, where it is temporarily excluded from convective mixing with the depleted mantle^{47–52}. The concentrations of Ce and Pb and the Ce/Pb ratio are likely to be high in the boundary layer because it would contain a higher percentage of relatively young oceanic crust than the convecting mantle. Isolation of this material would retard the increase of Ce/Pb and reduce the abundances of Ce and Pb in the MORB mantle, without increasing the Ce/Pb ratio of the continents. Thus a combination of convergent margin magmatism, boundary layer isolation of subducted crust and the additional continent-building mechanisms discussed above could produce appropriate Ce/Pb ratios in the continental crust and appropriate depletions of Ce and Pb in the MORB mantle.

Implications for mantle differentiation

Data from Umnak island show that fractionation of Pb from Ce by convergent margin processes is continuing. This apparently contradicts the formal three-stage mantle evolution model of Hofmann *et al.*⁵, which ignores the role of present-day mantle differentiation in fractionating the Ce/Pb ratio of the mantle. However, a broader interpretation of their model is consistent with our data. To explain the global Ce/Pb fractionation by current processes at convergent plate margins, oceanic crust must have been processed through subduction zones at a higher rate in the past, with a correspondingly higher rate of continental crust production. Subsequent reductions in the rate of crustal growth due to convergent margin magmatism would have also reduced the rate of Ce/Pb fractionation. Thus our data indicate that fractionation of Ce/Pb ratios in the mantle continues, but at a rate lower than the average over Earth history. □

Received 5 July 1993; accepted 1 March 1994.

- Jochum, K. P., Hofmann, A. W., Ito, E., Seufert, H. M. & White, W. M. *Nature* **306**, 431–436 (1983).
- Jochum, K. P., Hofmann, A. W. & Seufert, H. M. *Terra Cognita (Abstr.)* **6**, 186 (1986).
- Hofmann, A. W. *Earth Planet. Sci. Lett.* **90**, 297–314 (1988).
- Newsom, H. E., White, W. M., Jochum, K. P. & Hofmann, A. W. *Earth planet. Sci. Lett.* **80**, 299–313 (1986).
- Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. *Earth planet. Sci. Lett.* **79**, 33–45 (1986).
- Miller, D. M., Goldstein, S. L. & Langmuir, C. H. *EOS* **72**, 240 (1991).
- Hofmann, A. W. & Hart, S. R. *Terra Abstr.* **3**, 485 (1991).
- Chauvel, C., Goldstein, S. L. & Hofmann, A. W. *Terra Abstr.* **3**, 485 (1991).
- Peucker-Ehrenbrink, B., Hofmann, A. W. & Hart, S. R. *Earth Planet. Sci. Lett.* (submitted).
- Miller, D. M., Langmuir, C. H., Goldstein, S. L. & Franks, A. L. *J. geophys. Res.* **97**, 321–343 (1992).
- Miller, T. P. & Smith, R. L. *Geology* **15**, 434–438 (1987).
- Byers, F. M. Jr *Geology of Umnak and Bogoslof Islands, Aleutian Islands, Alaska* (US Geological Survey Bull. No. 1028-L, US Govt Printing Office, Washington DC, 1959).
- Perfit, M. R., Gust, D. A., Bence, A. E., Arculus, R. J. & Taylor, S. R. *Chem. Geol.* **30**, 227–256 (1980).
- Plank, T. & Langmuir, C. H. *Earth planet. Sci. Lett.* **90**, 349–370 (1988).
- Plank, T. & Langmuir, C. H. *Nature* **362**, 739–742 (1993).
- Kay, R. W., Sun, S. S. & Lee-Hui, C. N. *Geochim. cosmochim. Acta* **42**, 263–273 (1978).
- White, W. M. & Patchett, J. *Earth planet. Sci. Lett.* **67**, 167–185 (1984).
- Ben Othman, D., White, W. M. & Patchett, P. J. *Earth planet. Sci. Lett.* **94**, 1–21 (1989).
- McDermott, F. M. & Hawkesworth, C. *Earth planet. Sci. Lett.* **104**, 1–15 (1991).
- Stern, C. R. & Wyllie, P. J. *Am. Mineral.* **63**, 641–663 (1978).
- Gill, J. B. *Orogenic Andesites and Plate Tectonics* (Springer, New York, 1981).
- Dasch, E. J., Dymond, J. R. & Heath, G. R. *Earth planet. Sci. Lett.* **13**, 175–180 (1971).
- Goldstein, S. L. & O'Nions, R. K. *Nature* **292**, 324–327 (1981).
- Dasch, E. J. *Mem. geol. Soc. Am.* **54**, 199–209 (1981).
- LeHuray, A. P., Church, S. E., Koski, R. A. & Bouse, R. M. *Geology* **16**, 362–365 (1988).
- Davies, J. H. & Bickle, M. J. *Phil. Trans. R. Soc. A* **A335**, 355–364 (1991).
- Navon, O. & Stolper, E. J. *Geol.* **95**, 285–307 (1987).
- Karig, D. E. & Kay, R. W. *Phil. Trans. R. Soc. A* **301**, 233–251 (1981).
- Reymer, A. & Schubert, G. *Tectonics* **3**, 63–67 (1984).
- Sun, S.-s. & McDonough, W. F. in *Magmatism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313–345 (Spec. Publ. No. 42, Geological Soc., London, 1989).
- Galer, S. J. G., Goldstein, S. L. & O'Nions, R. K. *Chem. Geol.* **75**, 257–290 (1989).
- Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution* (Blackwell Scientific, Oxford, 1985).
- Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297–356 (1981).
- Cogley, J. G. *Rev. Geophys. Space Phys.* **22**, 101–122 (1984).
- Parsons, B. J. *geophys. Res.* **87**, 289–302 (1982).
- Allège, C. J., Hart, S. R. & Minster, J.-F. *Earth planet. Sci. Lett.* **66**, 191–213 (1983).
- Allège, C. J. & Lewin, E. *Earth planet. Sci. Lett.* **96**, 61–88 (1989).
- Galer, S. J. G. & O'Nions, R. K. *Nature* **316**, 778–782 (1985).
- Miller, D. M. Thesis, Columbia Univ. (in preparation).
- Nye, C. J. & Reid, M. R. *J. geophys. Res.* **91**, 10271–10287 (1986).
- von Drach, V., Marsh, B. D. & Wasserburg, G. J. *Contr. Miner. Petrol.* **92**, 13–34 (1986).
- Plank, T. Thesis, Columbia Univ. (1993).
- Furlong, K. P. & Fountain, D. M. *J. geophys. Res.* **91**, 8285–8294 (1986).
- Bohlen, S. R. & Mezger, K. *Science* **244**, 326–329 (1989).

45. Kay, R. W., Kay, S. M. & Arculus, R. J. in *Continental Lower Crust* (eds Fountain, D. M., Arculus, R. J. & Kay, R. W.) 423–445 (Elsevier, Amsterdam, 1992).
46. Abouchami, W., Boher, M., Michard, A. & Albarède F. *J. geophys. Res.* **95**, 17606–17629 (1990).
47. Morgan, W. J. *Nature* **230**, 42–43 (1971).
48. Chase, C. G. *Earth planet. Sci. Lett.* **52**, 277–284 (1981).
49. Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).
50. Christensen, U. *Annls Géophys.* **2**, 311–320 (1984).
51. Davies, G. F. *Earth planet. Sci. Lett.* **99**, 94–109 (1990).
52. Sleep, N. H. A. *Rev. Earth planet. Sci.* **20**, 19–43 (1992).
53. White, W. M. & Dupré, B. *J. geophys. Res.* **91**, 5927–5941 (1986).
54. Hémond, C. *Géochimie Isotopique du Th et du Sr dans la Série Tholéitique d'Islande et dans des Série Calco-Alcalines diverses* (Univ. Paris VII, France, 1986).
55. Kay, R. W. in *Island Arcs, Deep Sea Trenches and Back-Arc Basins* (eds Talwani, M. & Pitman, W. C.) 229–242 (Am. Geophys. Un., Washington DC, 1977).
56. Langmuir, C. H., Klein, E. M. & Plank, T. P. *Mantle Flow and Melt Generation at Mid-Ocean Ridges* 183–280 (AGU Monogr. No. 71, Am. Geophys. Un., Washington DC, 1992).
57. Chauvel, C., Hofmann, A. W. & Vidal, P. *Earth planet. Sci. Lett.* **110**, 99–119 (1992).
58. McCulloch, M. T. & Perfit, M. R. *Earth planet. Sci. Lett.* **56**, 167–180 (1981).
59. Oversby, V. M. *Geochim. cosmochim. Acta* **36**, 1167–1179 (1972).
60. Hawkesworth, C. J., O'Nions, R. K., Pankhurst, R. J., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **36**, 253–262 (1977).
61. Barreiro, B. *Geochim. cosmochim. Acta* **47**, 817–822 (1983).
62. Bardsell, M. & Berry, R. F. *J. Petrology* **31**, 747–777 (1990).
63. Thirwall, M. F. & Graham, A. M. *J. geol. Soc. Lond.* **141**, 427–445 (1984).

ACKNOWLEDGEMENTS. We thank C. Chauvel, S. Galer, T. Plank and J. Reynolds for discussion; B. Peucker-Ehrenbrink for a review and access to unpublished data; and A. Hofmann, M. Johnston, W. McDonough, H. Newsom, D. Walker and W. White for reviews. This work was partially supported by the US NSF.

Promoter targeting by adenovirus E1a through interaction with different cellular DNA-binding domains

Fang Liu^{*†} & Michael R. Green^{*}

[†] Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

^{*} Program in Molecular Medicine, University of Massachusetts Medical Center, 373 Plantation Street, Worcester, Massachusetts 01605, USA

A puzzling property of the transcriptional activators encoded by several animal viruses is their ability to function promiscuously. The adenovirus E1a protein, for example, stimulates transcription of adenoviral genes as well as a wide variety of other viral and cellular genes. We show that E1a can interact with several classes of cellular DNA-binding domains and thereby be recruited to diverse promoters. Our results explain how a single protein can regulate transcription of multiple genes that lack a common promoter element.

DURING the early phase of adenovirus (Ad) infection, two closely related E1a messenger RNAs, designated 13S and 12S on the basis of their sedimentation coefficients, are generated by alternative splicing. These two mRNAs encode 289- and 243-amino-acid E1a proteins, which differ by only 46 amino acids unique to the larger protein. This 46-amino-acid region, designated conserved region 3 (CR3), is necessary and sufficient for transcriptional activation of adenovirus early genes^{1–3}.

Protein-fusion experiments and mutational studies have shown that CR3 contains two separable activities^{4,5}. The N-terminal portion of CR3 contains a transcriptional activation region, functionally analogous to that of a typical cellular activator^{6,7}. The C-terminal portion of CR3 contains a 'promoter-targeting' region, required to direct the E1a activation region to its natural targets, the adenovirus early promoters⁴.

Because E1a is not a sequence-specific DNA-binding protein^{8,9}, how it targets the adenovirus early promoters has been unclear. One possibility is that E1a interacts with a cellular protein(s) that is bound to adenovirus early promoters. Several of these early promoters contain binding sites for the cellular activating transcription factors (ATFs), a family of proteins that contain homologous basic/leucine-zipper (bZIP) DNA-binding domains¹⁰. ATF sites have been implicated in E1a-inducible transcriptional activation³.

We and others have previously shown that a specific ATF, ATF-2, can mediate E1a-inducible transcriptional activation in a cotransfection assay; three other ATFs (ATF-1, ATF-4 and CREB) analysed to date, had little or no activity in the same assay (refs 11–13, and our unpublished results). Deletion analysis indicated that the N-terminal region of ATF-2, which

is absent from other ATFs, was essential for E1a responsiveness^{11–14}.

Here we present biochemical evidence for an interaction between E1a and ATF-2. Surprisingly, the portion of ATF-2 required for this interaction is not the N-terminal region, but rather the bZIP DNA-binding domain. Remarkably, we find that E1a can also interact with unrelated DNA-binding domains of several cellular transcription factors. Our results help explain how E1a can activate transcription of a variety of viral and cellular genes whose promoters differ significantly¹⁵.

E1a interacts with ATF-2 *in vitro*

To determine whether E1a and ATF-2 interact, we performed *in vitro* protein-affinity chromatography experiments. We expressed in *Escherichia coli* a glutathione S-transferase (GST)–ATF2 fusion protein and analysed its ability to bind ³⁵S-labelled 12S and 13S E1a products synthesized by *in vitro* translation. As a positive control for E1a binding, we used a previously described fusion protein of GST with the retinoblastoma protein Rb¹⁶. The ³⁵S-labelled E1a proteins were incubated with glutathione-agarose beads containing immobilized GST, GST–ATF2 or GST–Rb. After extensive washing, bound proteins were fractionated on an SDS-polyacrylamide gel and visualized by autoradiography. Figure 1a shows that neither 13S nor 12S E1a bound to the control GST protein. The 13S E1a product, which contains CR3, bound to GST–ATF2, whereas 12S E1a, which lacks CR3, did not. Both 13S and 12S E1a bound efficiently to GST–Rb (Fig. 1a), consistent with the fact that E1a interacts with Rb through CR1 and CR2 (ref. 17), which are present in both E1a products. Both the E1a–Rb and

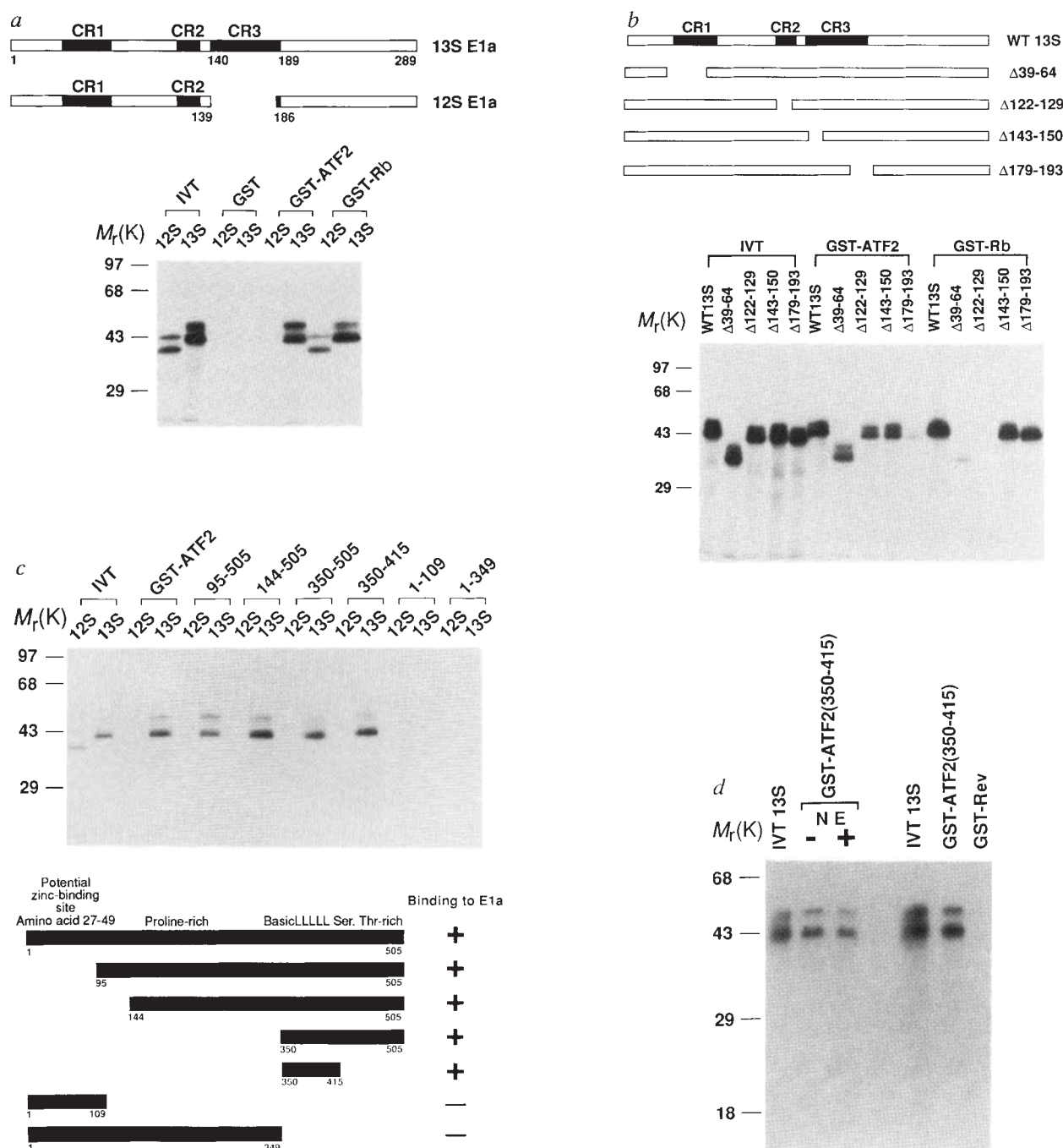


FIG. 1 E1a interacts with the bZIP DNA-binding domain of ATF-2. *a*, ^{35}S -labelled 12S and 13S E1a products were synthesized by *in vitro* translation. E1a protein (10 μl) was diluted in 1 ml binding buffer and incubated with glutathione-agarose beads containing $\sim 35 \mu\text{g}$ GST, $\sim 30 \mu\text{g}$ GST-ATF2 or $\sim 0.5 \mu\text{g}$ GST-Rb. Half of the bound proteins were analysed on a 10% SDS-polyacrylamide gel. The first two lanes show 1 μl *in vitro*-translated (IVT) 12S and 13S E1a products, respectively. The structures of the 12S and 13S E1a products are shown above the autoradiogram. CR1, CR2 and CR3 refer to conserved regions 1, 2 and 3 of E1a proteins, respectively². *b*, E1a mutants were analysed for binding to ATF-2 and Rb as described in *a*. *c*, E1a proteins (12S and 13S) were incubated with glutathione-agarose beads containing $\sim 30 \mu\text{g}$ GST-ATF2, $\sim 25 \mu\text{g}$ GST-ATF2 (95–505), $\sim 20 \mu\text{g}$ GST-ATF2 (144–505), $\sim 20 \mu\text{g}$ GST-ATF2 (350–505), $\sim 10 \mu\text{g}$ GST-ATF2 (350–415), $\sim 45 \mu\text{g}$ GST-ATF2 (1–109) or $\sim 35 \mu\text{g}$ GST-ATF2 (1–349). The diagram below the autoradiogram summarizes the results. *d*, 13S E1a product was incubated with glutathione-agarose beads containing $\sim 1.6 \mu\text{g}$ GST-ATF2 (350–415) in the absence or presence of 75 μg

HeLa cell nuclear extract (NE; left), or incubated with $\sim 2 \mu\text{g}$ GST-ATF2 (350–415) or GST-Rev (right).

METHODS. GST-ATF2 and derivatives were constructed by fusing an appropriate DNA fragment in-frame to the GST sequence. GST-Rev (a gift from Maria L. Zapp) contains full-length Rev (amino acids 1–116) fused to GST. GST-Rb, SP12SE1a, SP13SE1a and SPE1a (Δ 39–64) (SP5/3) have been described^{16,40}. E1a (Δ 122–129) was derived from dI922/947 (ref. 17). E1a (Δ 143–150) and E1a (Δ 179–193) were derived by *Bal*/31 deletion of SP13SE1a. The GST fusion proteins were induced and purified essentially as described^{16,41}. ^{35}S -labelled E1a synthesized in the presence of 100 μM ZnCl_2 and 20 mM β -mercaptoethanol was incubated with $\sim 35 \mu\text{l}$ glutathione-agarose beads containing a GST derivative on a rocker at 4 $^\circ\text{C}$ for 2 h. Bound proteins were analysed after 5 washes each with 1 ml buffer for 10 min on ice. The composition of binding and washing buffer was: 40 mM HEPES, pH 7.4, 100 mM KCl, 100 μM ZnCl_2 , 0.1% N-P40, 20 mM β -mercaptoethanol, 0.1 mM PMSF.

E1a-ATF2 interactions were unaffected by 100 $\mu\text{g ml}^{-1}$ ethidium bromide (data not shown), indicating that these interactions were not due to contaminating DNA¹⁸. We conclude that E1a binds to ATF-2 *in vitro* and that CR3 is required for this interaction.

E1a promoter-targeting region interacts with ATF-2

We next sought to delineate the precise region of E1a required for interaction with ATF-2. To determine whether sequences outside CR3 were involved, we analysed two in-frame deletion mutants (Fig. 1b). A CR1 deletion, E1a ($\Delta 39-64$), and a CR2 deletion, E1a ($\Delta 122-129$), bound to GST-ATF2 at levels comparable to that of wild-type E1a. In contrast, binding of both E1a mutants to GST-Rb was severely reduced. To define the portion of CR3 required for binding to ATF-2, we tested two deletion mutants. E1a ($\Delta 143-150$), which lacks the N-terminal transcriptional activation region of CR3, bound efficiently to both GST-ATF2 and GST-Rb (Fig. 1b). In contrast, E1a ($\Delta 179-193$), which bears a deletion of 15 amino acids (CGMFVYSPVSEPEPE) in the CR3 promoter-targeting region, bound very poorly to GST-ATF2 but bound like wild-type E1a to GST-Rb (Fig. 1b). On the basis of these results, we conclude that E1a interacts with ATF-2 *in vitro* via a portion of CR3 previously implicated in promoter targeting through *in vivo* experiments^{4,11}.

E1a interacts with ATF-2 bZIP DNA-binding domain

The N-terminal 143 amino acids of ATF-2 is required for the E1a transcriptional response *in vivo*¹¹⁻¹⁴. To determine whether this is the E1a binding site, we analysed a series of GST-ATF2 derivatives (Fig. 1c). To our surprise, removal of the N-terminal 94 amino acids or 143 amino acids of ATF-2 did not affect binding to E1a. Furthermore, GST fusion proteins containing only the N-terminal 109 or 349 amino acids of ATF-2 did not bind E1a. Analysis of additional N-terminal and C-terminal deletion mutants delineated 65 amino acids (350-415) comprising the ATF-2 bZIP DNA-binding domain as sufficient for E1a binding.

Figure 1d presents two additional control experiments to confirm the specificity of the E1a-bZIP interaction. First, addition of 75 μg of HeLa cell nuclear extract did not prevent binding of E1a. Second, because the bZIP domain is positively charged, it was important to ensure that E1a binding was not due to a nonspecific ionic interaction. Figure 1d shows that E1a did not bind to the human immunodeficiency virus Rev protein, which contains a highly positively charged RNA-binding domain¹⁹.

E1a interacts with multiple ATFs

The bZIP DNA-binding domain is highly conserved among ATF members¹⁰. The finding that E1a interacts with the bZIP DNA-binding domain of ATF-2 raised the possibility that E1a could also interact with other ATFs. Figure 2a shows that 13S E1a, but not 12S E1a, bound to ATF-1 and ATF-3 at a level comparable to that of ATF-2.

N-terminal region of ATF-2 and E1a response

The finding that E1a interacts with multiple ATFs presents a paradox: why does ATF-2, but not other ATFs, support an E1a response^{11,13}? Because the ATF-2 N terminus has been previously implicated in E1a responsiveness¹¹, we next investigated how this region functions. We first verified that the failure of other ATFs to support an E1a response was indeed due to the lack of such an N-terminal region. We fused the N-terminal 109 amino acids of ATF-2 to ATF-1, and this ATF2(N)-ATF1 chimera was then fused to the GAL4 DNA-binding domain. Figure 2b shows that GAL4-ATF2(N)-ATF1, but not the control GAL4-ATF1, supported transcriptional activation by E1a.

The N-terminal region of ATF-2 is necessary for E1a responsiveness, but it does not interact with E1a. What then is the activity provided by this region? We tested the possibility that the

N-terminal region of ATF-2 contains a transcriptional activation domain. Figure 3a shows that a GAL4 fusion protein containing ATF-2 amino acids 1-109 activated transcription from a promoter containing GAL4-binding sites (see also refs 13, 14). Cotransfection with E1a did not significantly increase the activity of GAL4-ATF2 (1-109) (data not shown), presumably because GAL4-ATF2 (1-109) lacks the E1a-binding region. Consistent with our previous results, a GAL4 fusion protein containing full-length ATF-2 failed to stimulate transcription (Fig. 3a). Thus, the N-terminus of ATF-2 contains an activation region, which is 'cryptic' in the intact protein. Figure 3b compares the strength of the ATF-2 activation region to that of several well characterized activation domains. We conclude that the N terminus of ATF-2 contains a transcriptional activation region required for E1a responsiveness.

E1a sequences outside CR3 inhibit activation

The observation that E1a function requires the ATF-2 activation region was unexpected in the light of earlier results. For example,

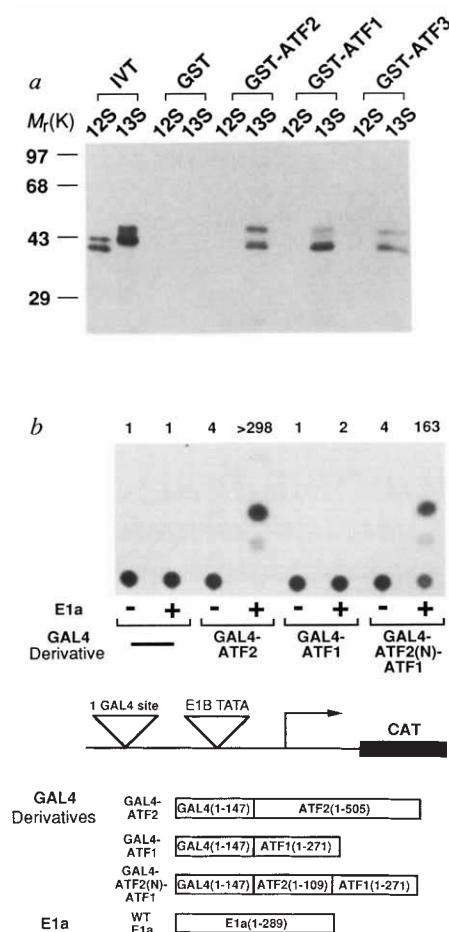


FIG. 2 E1a interacts with other ATFs. *a*, *In vitro*: E1a proteins 12S and 13S were incubated with glutathione-agarose beads containing $\sim 30 \mu\text{g}$ GST-ATF2, $\sim 20 \mu\text{g}$ GST-ATF1, or $\sim 20 \mu\text{g}$ GST-ATF3. *b*, *In vivo*: CHO cells were cotransfected with a CAT reporter (4 μg DNA), expression plasmids encoding GAL4 derivatives (700 ng DNA) and E1a (4 μg DNA) as indicated. CAT activity was quantified by scintillation counting and is indicated above each lane.

METHODS. *a*, GST-ATF1, GST-ATF3, GAL4-ATF2 and GAL4-ATF1 have been described^{11,20}. GAL4-ATF2 (1-109)-ATF1 (1-271) was constructed by fusing the appropriate DNA fragment in-frame to the GAL4 (1-147) sequence in the vector pSG424 (ref. 42). Transfection and CAT assays were carried out as described¹¹.

GAL4-E1a (121–223), which contains CR3, can efficiently activate transcription in the absence of ATF-2 (ref. 4). One explanation for this difference could be that full-length E1a, like ATF-2, contains sequences that inhibit its minimal-activation domain. To test this idea, we compared the activity of GAL4-E1a (121–223) to that of a GAL4 fusion protein containing full-length E1a, GAL4-E1a (1–289). Figure 4a shows that GAL4-E1a (121–223) efficiently activated transcription, whereas

GAL4-E1a (1–289) was a very poor activator. The weak activity of full-length E1a explains why efficient activation requires cooperation with another activation domain.

In our previous study, E1a activated transcription when targeted to the promoter through a GAL4-Rb fusion protein¹¹. In contrast, E1a failed to activate transcription when targeted through several GAL4-bZIP fusion proteins such as GAL4-ATF1 (Fig. 2b; see also ref. 11) and GAL4-ATF2 (350–505)¹¹. A possible explanation for this apparent difference is that the Rb protein, which interacts with E1a through CR1 and CR2, may counteract the negative effect of sequences flanking CR3. Consistent with this idea, Fig. 4b shows that cotransfection of a Rb expression plasmid increases the activity of GAL4-E1a (1–289). An alternative explanation is that Rb contains a weak activation region that can cooperate with that of E1a. In this regard, Rb can stimulate transcription in cooperation with several cellular transcription factors such as ATF-2 (ref. 20) and SP1 (refs 21–22).

E1a interacts with different DNA-binding domains

E1a stimulates transcription of more than 20 heterologous viral and cellular genes¹⁵. Our results prompted us to ask whether E1a can also interact with the DNA-binding domains of other cellular transcription factors. We constructed GST fusion proteins containing the bZIP DNA-binding domain of c-Jun, the zinc-finger DNA-binding domain of Sp1, and the basic/helix-loop-helix (bHLH) DNA-binding domain of USF. Figure 5 shows that 13S E1a, but not 12S E1a, bound to these three GST fusion proteins, but did not bind to the DNA-binding domain of the yeast GAL4 protein or GST (data not shown). As with ATF-2 (Fig. 1b), deletion within the E1a promoter targeting region (E1a (Δ 179–193)) significantly reduced the association with each of these DNA-binding domains (Fig. 5). Ethidium bromide (100 μ g ml⁻¹) did not disrupt any of these interactions, indicating that contaminating DNA was not involved¹⁸.

To demonstrate that these *in vitro* interactions were functionally relevant, we analysed the ability of GAL4 fusion proteins containing c-Jun, Sp1 or USF to support E1a-mediated activation in a cotransfection assay. Figure 6a shows that GAL4-cJun

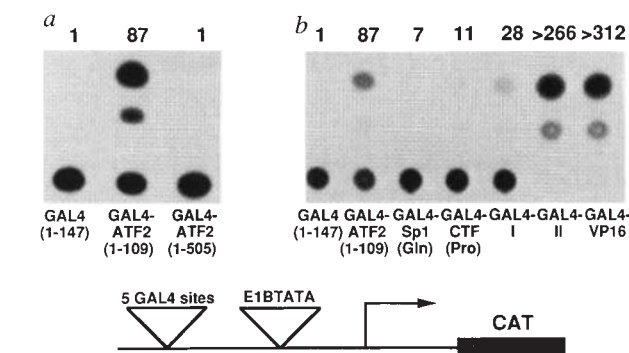


FIG. 3 The N-terminal region of ATF-2 contains a transcriptional activation region. **a**, GAL4-ATF2 (1–109) stimulates transcription from a GAL4 reporter. **b**, Comparison of transcriptional activity of GAL4-ATF2 (1–109) with those of several well characterized activators. CAT activity was quantified by scintillation counting and is indicated above each lane.

METHODS. GAL4-Sp1 and GAL4-CTF (Pro) have been described^{43,44}. CHO cells were cotransfected with a CAT reporter (4 μ g DNA) and expression plasmids encoding GAL4 derivatives. For each activator, a titration was used to determine the amount of expression plasmid giving rise to maximal activity.

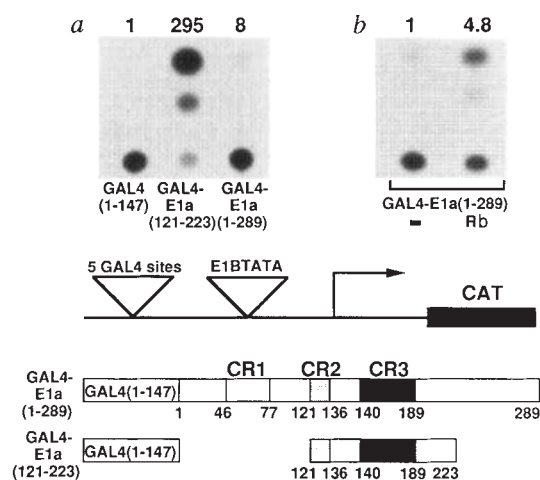


FIG. 4 E1a sequences outside CR3 inhibit transcription. **a**, Comparison of GAL4-E1a (1–289) and GAL4-E1a (121–223). CHO cells were cotransfected with a CAT reporter (4 μ g DNA) and GAL4-E1a (1–289) or GAL4-E1a (121–223). For each activator, a titration was done by varying the amount of expression plasmid. 1 μ g GAL4-E1a (121–223), 200 ng GAL4-E1a (1–289) and 2 μ g GAL4 (1–147) gave maximal activity and are shown here. **b**, The p105-Rb protein increases transcriptional activity of GAL4-E1a (1–289). CHO cells were cotransfected with a CAT reporter (4 μ g DNA), GAL4-E1a (1–289) (200 ng DNA) and pECE-Rb (an Rb expression plasmid; 4 μ g DNA). CAT activity was quantified by scintillation counting and is indicated above each lane.

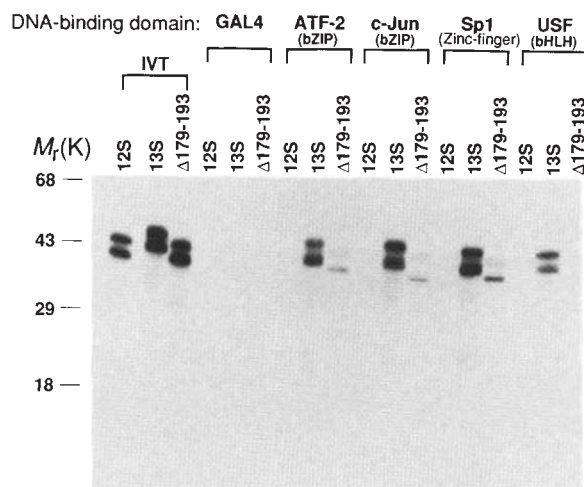


FIG. 5 E1a can interact with different DNA-binding domains *in vitro*. E1a proteins were incubated with glutathione-agarose beads containing ~12 μ g GST-GAL4 (1–147), ~7 μ g GST-ATF2 (350–415), ~14 μ g GST-cJun (222–331), ~2.5 μ g GST-Sp1 (612–768), or ~15 μ g GST-USF (194–262). Identical results were obtained when Mg²⁺ was included in the binding buffer (data not shown).

efficiently supported E1a inducibility, whereas GAL4-Sp1 and GAL4-USF only moderately supported an E1a response.

Based upon our results with ATF proteins, we considered that the low transcription with Sp1 and USF might be due to the lack of an appropriate activation region. To test this possibility,

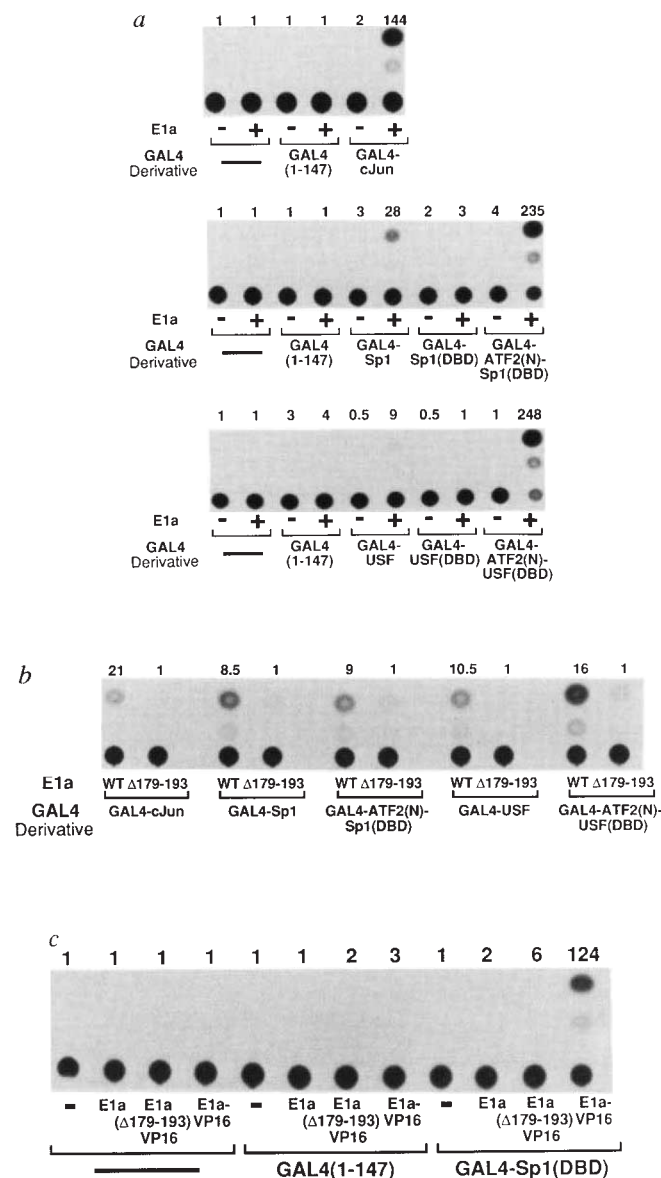


FIG. 6 E1a can interact with different DNA-binding domains *in vivo*. **a**, CHO cells were cotransfected with a CAT reporter (4 μ g DNA), GAL4 derivatives (top and middle, 200 ng DNA; bottom, 2 μ g DNA) and E1a (4 μ g DNA). **b**, CHO cells were cotransfected with a CAT reporter (4 μ g DNA), GAL4 derivatives (200 ng DNA for GAL4-cJun and GAL4-Sp1, 40 ng for GAL4-ATF2(N)-Sp1(DBD), 2 μ g for GAL4-USF, and 500 ng for GAL4-ATF2(N)-USF(DBD)), and wild-type (WT) 13S E1a or 13S E1a (Δ 179–193) mutant (4 μ g DNA for GAL4-cJun, GAL4-Sp1 and GAL4-USF, 1 μ g DNA for GAL4-ATF2(N)-Sp1(DBD) and GAL4-ATF2(N)-USF(DBD)). **c**, CHO cells were cotransfected with a CAT reporter (4 μ g DNA), GAL4 derivatives (200 ng DNA) and E1a, E1a (Δ 179–193)-VP16 or E1a-VP16 (1 μ g DNA). **METHODS**. GAL4-cJun (1–331), GAL4-Sp1(DBD) (612–778), GAL4 ATF2 (1–109)-Sp1 (612–778), GAL4-USF (1–310), GAL4-USF(DBD) (194–262), and GAL4-ATF2 (1–109)-USF (194–262) were constructed by fusing the appropriate DNA fragment in-frame to the GAL4 (1–147) sequence in the vector pSG424 (ref. 42). E1a (Δ 179–193)-VP16 and E1a-VP16 contains E1a (1–289 (Δ 179–193)) and E1a (1–289) fused to VP16 (413–490), respectively.

we did two types of experiments. First, we fused the N-terminal region of ATF2 (amino acids 1–109) to the DNA-binding domains (DBD) of Sp1 and USF. Figure 6a shows that GAL4-ATF2(N)-Sp1 (DBD) and GAL4-ATF2(N)-USF (DBD) efficiently supported an E1a response. Figure 6b shows that, in each case, an in-frame deletion of the E1a promoter-targeting region (Δ 179–193) inhibited activation.

Second, we fused to E1a the potent acidic activation region of the herpes simplex virus (HSV) VP16 protein. Figure 6c shows that a GAL4 fusion protein containing the SP1 DNA-binding domain supported transcriptional activation by E1a-VP16, but not by E1a (Δ 179–193)-VP16. In this experiment E1a-VP16 is recruited to the promoter by an interaction between E1a and the promoter-bound Sp1 DNA-binding domain.

Discussion

On the basis of our previous studies and the results presented here, we propose a model for specific and promiscuous activation by E1a. E1a targets a promoter by interacting with the DNA-binding domain of a cellular transcription factor. The activation regions of E1a and the cellular protein then cooperate to stimulate transcription.

For several adenovirus early promoters, it is the ATF-2 protein with which E1a interacts. E1a can also stimulate transcription of a wide variety of other viral and cellular promoters. We have shown that E1a interacts with diverse DNA-binding domains including those of c-Jun, Sp1 and USF. The relevance of these *in vitro* interactions is supported by two lines of evidence. First, our protein-fusion experiments demonstrate that these three cellular transcription factors can mediate an E1a response *in vivo*. Second, previous studies have shown that AP-1, Sp1 and USF sites can all confer E1a inducibility^{23,30}. The finding that E1a can interact with diverse DNA-binding domains explains, at least in part, its ability to stimulate transcription promiscuously.

Synergy with cellular transcription factors

Although E1a can interact with a variety of transcription factors, efficient transcription requires, in addition, synergy with the cellular DNA-binding protein. E1a displays varying abilities to cooperate with different cellular activators: E1a cooperates efficiently with ATF-2 and c-Jun; and poorly (or not at all) with Sp1, USF and several acidic activators (Fig. 6a and ref. 11). Dramatic variations in the abilities of different combinations of cellular activators to cooperate have been previously reported^{31,32}.

The basis for transcriptional synergy between eukaryotic activators is unknown. One possibility is that the different activators, such as E1a and ATF-2, interact with distinct components of the general transcription machinery. Transcriptional synergy could also involve a conformational change in ATF-2, E1a, or both, which appropriately presents the activation domains to the general transcription machinery.

Multiple functions of DNA-binding domains

We have shown that E1a interacts with cellular transcription factors through association with their DNA-binding domains. The structural basis for this interaction of E1a with apparently unrelated DNA-binding domains remains to be elucidated. Our findings are in accord with observations that DNA-binding domains have additional regulatory functions^{33,38}. DNA-binding domains are generally the most conserved portions of eukaryotic transcription factors. By targeting this region, E1a, and perhaps other viral activators, can act upon many viral and cellular genes. Certain cellular regulators, such as SWI/SNF proteins, also modulate transcription of a wide variety of genes³⁹. In fact, one such protein, SWI3, interacts with the glucocorticoid receptor through its DNA-binding domain³⁶. These observations raise the possibility that other promiscuous activators may work by a mechanism similar to that used by E1a. □

Received 14 December 1993; accepted 8 February 1994.

1. Berk, A. J. *Rev. Genet.* **20**, 45–79 (1986).
2. Moran, E. & Mathews, M. B. *Cell* **48**, 177–178 (1987).
3. Flint, J. & Shenk, T. A. *Rev. Genet.* **23**, 141–161 (1989).
4. Lillie, J. W. & Green, M. R. *Nature* **338**, 39–44 (1989).
5. Martin, K. J., Lillie, J. W. & Green, M. R. *Nature* **346**, 147–152 (1990).
6. Ptashne, M. *Nature* **335**, 683–689 (1988).
7. Mitchell, P. J. & Tjian, R. *Science* **245**, 371–378 (1989).
8. Ferguson, B., Krippel, B., Andrisani, O., Jones, N. & Westphal, H. *Molec. cell. Biol.* **5**, 2653–2661 (1985).
9. Chatterjee, P. K., Bruner, M., Flint, S. J. & Harter, M. L. *EMBO J.* **7**, 835–841 (1988).
10. Hai, T., Liu, F., Coukos, W. J. & Green, M. R. *Genes Dev.* **3**, 2083–2090 (1989).
11. Liu, F. & Green, M. R. *Cell* **61**, 1217–1224 (1990).
12. Maekawa, T., Matsuda, S., Fujisawa, J., Yoshida, M. & Ishii, S. *Oncogene* **6**, 627–632 (1991).
13. Flint, K. J. & Jones, N. C. *Oncogene* **6**, 2019–2026 (1991).
14. Zu, Y.-L., Maekawa, T., Matsuda, S. & Ishii, S. *J. biol. Chem.* **266**, 24134–24139 (1991).
15. Shenk, T. & Flint, J. *Adv. Cancer Res.* **57**, 47–85 (1991).
16. Kaelin, W. G., Pallas, D. C., DeCaprio, J. A., Kaye, F. J. & Livingston, D. M. *Cell* **64**, 521–532 (1991).
17. Whyte, P., Williamson, N. M. & Harlow, E. *Cell* **56**, 67–75 (1989).
18. Lai, J.-S. & Herr, W. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6958–6962 (1992).
19. Tao, J. & Frankel, A. D. in *AIDS Res. Rev.* 3–22 (1993).
20. Kim, S.-J. et al. *Nature* **358**, 331–334 (1992).
21. Kim, S.-J. et al. *Molec. cell. Biol.* **12**, 176–183 (1992).
22. Udvardi, A. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3265–3269 (1993).
23. Lewis, E. D. & Manley, J. L. *Molec. cell. Biol.* **5**, 1044–1051 (1985).
24. Hurst, H. C. & Jones, N. C. *Genes Dev.* **1**, 1132–1146 (1987).
25. Buckbinder, L., Miralles, V. J. & Reinberg, D. *EMBO J.* **8**, 4239–4250 (1989).
26. Chang, L.-S., Shi, Y. & Shenk, T. *J. Virol.* **63**, 3479–3488 (1989).
27. Taylor, I. C. A. & Kingston, R. *Molec. cell. Biol.* **10**, 176–183 (1990).
28. van Dam, H. et al. *Molec. cell. Biol.* **10**, 5857–5864 (1990).
29. de Groot, R., Foulkes, N., Mulder, M., Kruijer, W. & Sassone-Corsi, P. *Molec. cell. Biol.* **11**, 192–201 (1991).
30. Kitabayashi, I., Chiu, R., Gachelin, G. & Yokoyama, K. *Nucleic Acids Res.* **19**, 649–655 (1991).
31. Lin, Y.-S., Carey, M. F., Ptashne, M. & Green, M. R. *Cell* **54**, 659–664 (1988).
32. Wang, W. & Gralla, J. D. *Molec. cell. Biol.* **11**, 4561–4571 (1991).
33. Stern, S., Tanaka, M. & Herr, W. *Nature* **341**, 624–630 (1989).
34. Miner, J. N. & Yamamoto, K. R. *Trends Biochem. Sci.* **16**, 423–426 (1991).
35. Davis, R. L. & Weintraub, H. *Science* **256**, 1027–1030 (1992).
36. Yoshinaga, S. K., Peterson, C. L., Herskowitz, I. & Yamamoto, K. R. *Science* **258**, 1598–1604 (1992).
37. Du, W., Thanos, D. & Maniatis, T. *Nature* **74**, 887–898 (1993).
38. Wagner, S. & Green, M. R. *Science* **262**, 395–399 (1993).
39. Winston, F. & Carlson, M. *Trends Genet.* **8**, 387–391 (1992).
40. Richter, J. D., Slavicek, J. M., Schneider, J. F. & Jones, N. C. *J. Virol.* **62**, 1948–1955 (1988).
41. Lin, Y.-S. & Green, M. R. *Cell* **64**, 971–981 (1991).
42. Sadowski, I. & Ptashne, M. *Nucleic Acid Res.* **17**, 7539 (1989).
43. Southgate, C. D. & Green, M. R. *Genes Dev.* **5**, 2496–2507 (1991).
44. Tanese, N., Pugh, B. F. & Tjian, R. *Genes Dev.* **5**, 2212–2224 (1991).

ACKNOWLEDGEMENTS. We thank B. Kaelin and D. Livingston for the GST-Rb construct; N. Jones and N. LaThangue for 12SE1a, 13SE1a and E1a ($\Delta 39$ –64) *in vitro* expression plasmids; T. Deng and M. Karin for *c-jun* plasmids; G. Gill and R. Tjian for GAL4-Sp1 plasmid; H. Du, K. Chueng and R. Roeder for USF plasmids; W. Lee and A. Berk, J. Geigsberg and R. Ricciardi, L. Webster and A. Lassar, M. Zapp and X.-Y. Li for constructs; and J. Lillie, S. Roberts, X.-Y. Li and other members of our laboratory for critically reading the manuscript and for discussion. This work was supported by a grant from the NIH to M.R.G.

LETTERS TO NATURE

Dynamic response of Jupiter's atmosphere to the impact of comet Shoemaker–Levy 9

Joseph Harrington, Raymond P. LeBeau Jr, Karl A. Backes & Timothy E. Dowling

Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

DURING the period 18–24 July 1994, over 20 fragments of comet Shoemaker–Levy 9 will collide with Jupiter^{1–3}. The thermal and condensation signatures of inertia-gravity waves emanating from the impact sites will, if detectable, provide valuable insight into the stratification of Jupiter's atmosphere. We report here simulations of the event using a global multi-layer model⁴ of Jupiter's atmosphere and a range of impact kinetic energies (10^{27} – 10^{30} erg) that allows for the uncertainties in the sizes and densities of the comet fragments^{5–8}. The resulting inertia-gravity waves give rise to temperature perturbations in the range 0.004–1.2 K. The signature of the larger impacts may be detectable by thermal infrared imaging, and even weak signals may be detectable if one allows for the fact that the waves propagate in coherent rings centred on each impact site. Our simulations also indicate that a small vortex should form in the atmosphere following each impact, but that these will be sheared apart by the zonal winds within a few weeks.

An atmosphere's density stratification, the rate at which density decreases with altitude, strongly influences the type of weather it exhibits. On Jupiter, stratification is poorly constrained in the most active region of the atmosphere, the troposphere. The deformation radius, L_d , is one measure of stratification. For length scales smaller than L_d , gravity flattens pressure highs and lows; for length scales larger than L_d , the Coriolis force sustains these anomalies. On Jupiter, L_d ranges from ~3,000 km in the stratosphere⁹ to zero in the neutrally-stable convecting interior. Observational data poorly constrain the transition between these extremes, yet the transition greatly affects tropospheric meteorology. For example, the effective L_d in two-layer atmosphere models^{10–13} is uncertain by a factor of five, and as L_d is usually squared in fluid equations, key dynamical terms are uncertain by a factor of 25.

One can determine L_d by measuring the speed of inertia-gravity waves. These waves are generated by the adjustment process that brings large-scale disturbances into geostrophic balance. The group velocity of the leading wavefront is $c \approx L_d f$, where $f = 2\Omega \sin(\lambda)$ is the Coriolis parameter, Ω is the planetary rotation rate, and λ is the planetographic latitude. The Shoemaker–Levy 9 impacts may provide the perturbations needed to set observable waves in motion, after which wave speeds will not depend on anything to do with the comet—a perfect experiment and one that is unlikely to recur soon¹⁴.

To investigate the dynamical response to comet impacts, we ran simulations with the Explicit Planetary Isentropic-Coordinate (EPIC) atmospheric model for Jupiter⁴, which is based on the finite-difference algorithm of Hsu and Arakawa¹⁵. To resolve Jupiter's nonlinear dynamics, our horizontal resolution is 512×256 cells, equivalent to 0.7° or ~900 km at the equator. Five active layers simulate the atmosphere (see Fig. 1), and a sixth layer with a steady wind profile models the interior. Specifying the initial winds in each layer is problematical¹⁶. The Voyager infrared observations and the thermal wind equation¹⁷ indicate that Jupiter's winds decay with height above the cloud tops. How they vary below the clouds is unknown, although we expect the deep interior to rotate with the magnetic field. In the simulations presented here, we constructed the initial wind field in each active layer by multiplying the observed cloud-top zonal wind¹⁸ by a gaussian function in altitude centred at 500 mbar with a width parameter $\sigma = 2$ pressure scale heights. We set the interior profile to half the cloud-top winds. The choice of interior wind affects planetary-scale (Rossby) waves but does not significantly affect inertia-gravity waves, which are the focus of this work. We gradually force the initial winds for ~50 d to allow the mass and momentum fields to balance. The model then runs without forcing for ~150 d so shear instabilities can develop and equilibrate before the impact perturbation.

The EPIC model was designed for large-scale meteorological applications and hence incorporates the hydrostatic approximation, which equates the vertical pressure gradient with gravity and ignores vertical accelerations. This unfortunately eliminates sound waves that, in the Shoemaker–Levy 9 event, may provide seismic information about Jupiter's deep interior¹⁹. EPIC employs a wave-damping 'sponge', gradually introduced in the upper 20% of the layers to prevent reflections from the top of the model²⁰. For the runs reported here only the top layer contains the sponge, but it is nevertheless effective; future work will

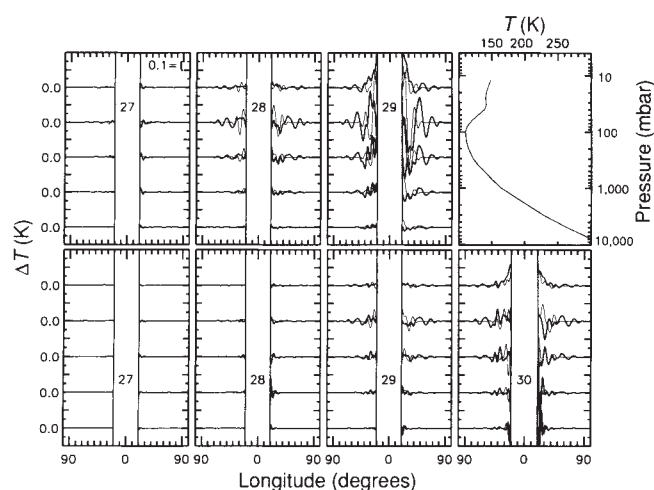


FIG. 1 Upper right, The initial temperature versus pressure, $T(p)$, curve for the model atmosphere, showing the tropopause at ~ 100 mbar. We used the available Voyager data for pressures less than 700 mbar. For higher pressures we extrapolated as shown assuming a slightly stable atmosphere. Such guess-work will be unnecessary if inertia-gravity wave speeds are measured in the troposphere. Other plots, Temperature deviation profiles for comet-Jupiter impact simulations. The top row of three plots corresponds to energy deposition into the stratosphere, the lower row of four plots to tropospheric energy deposition. The five vertically-stacked profiles in each of the seven plots correspond to layer interfaces in the model. The top layer extends to zero pressure and the steady-wind interior begins at 5,000 mbar. The plots show temperature deviations from an unperturbed model one day (light line) and two days (heavy line) after the impact of a comet fragment. Minor ticks on the vertical axis are 0.1 K. Energy deposition altitudes are indicated by the placement of the log of the impact energy in ergs. To determine interface pressure levels (and hence layer spacings), extend the zero of the vertical scale to the pressure axis of the $T(p)$ plot. The region close to the impact point is masked for clarity. A low-pass filter removed grid-scale noise.

use more layers. A low-pass filter at the poles prevents numerical instabilities resulting from the small grid spacing.

The impact velocity is well-established³ at $\sim 60 \text{ km s}^{-1}$. On the other hand, disparate estimates of the maximum fragment size (1–4 km, refs 5 and 6) and varying density assumptions (0.2 – 1.0 g cm^{-3} , refs 7 and 8) combine to make the impact kinetic energy uncertain by a factor of ~ 400 . Further, no reliable constraint has been placed on the energy fraction that remains in the atmosphere after the nonhydrostatic phase. Models^{7,8,21} of the first minutes after an impact vary greatly in their predictions. Reference 21 predicts fragment penetration to 100 bar and a gradual release of energy, ref. 7 predicts explosive vaporization at 10–100 mbar, whereas ref. 8 predicts it at 10 bar. Reference 8 further predicts a rapid ($\sim 10 \text{ km s}^{-1}$) rise of superheated gas that escapes the atmosphere and then falls back down onto the stratosphere over a range of several thousand kilometres.

Because there are ~ 20 fragments of various brightnesses, we expect a range of actual impact altitudes and energies.

To bracket these uncertainties, we modelled heat deposition in the range of 10^{27} – 10^{30} erg at the predicted impact latitude of 43.9° S (10^{28} erg corresponds to a 1-km diameter fragment of density 1 g cm^{-3} and mass $5 \times 10^{14} \text{ g}$). Our hydrostatic model cannot handle large vertical accelerations, so we start our simulations well after the rise of gas by adding heat to a single layer instantaneously. The heat is spread over as small a disc as possible without violating the hydrostatic assumption; diameters are a few thousand kilometres. We did not run a 10^{30} -erg stratospheric case as the disc size was unrealistically large. Because models of the early event disagree on where the energy will ultimately go, we simulated stratospheric and tropospheric depositions, in the layers spanning 16–69 mbar and 287–1,197 mbar, respectively.

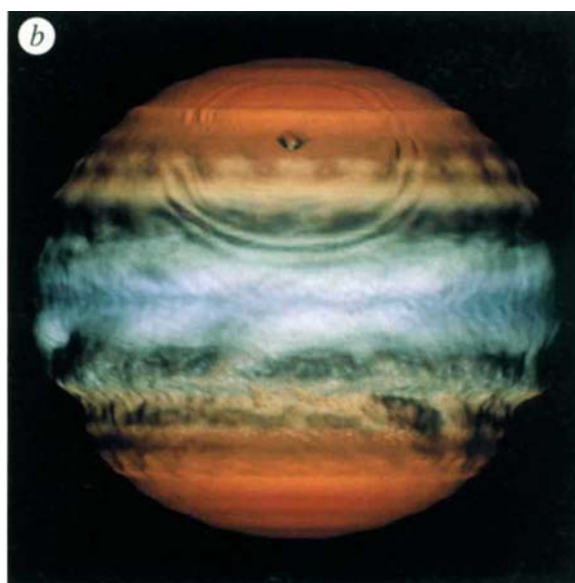
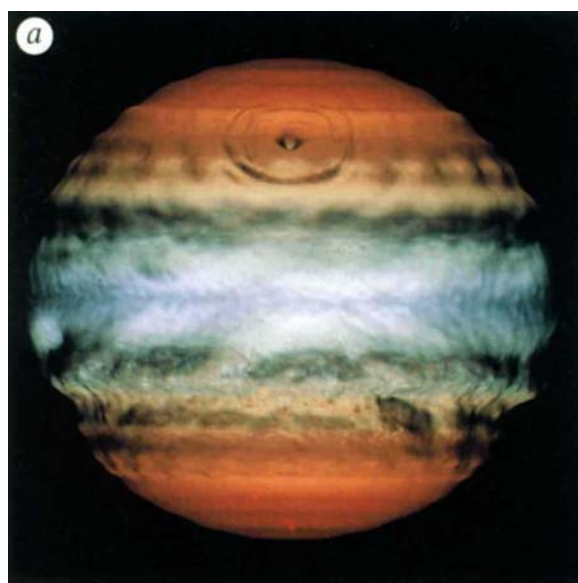


FIG. 2 a, Simulation of Jupiter's atmosphere one day (24 h) after impact of a comet fragment. b, Same simulation two days after impact. 10^{28} erg were deposited in the layer spanning 16–69 mbar (stratospheric deposition); the 69-mbar pressure level is shown here. Height variations on the spheres correspond to changes in pressure, whereas colour corresponds to potential vorticity and indicates material initially at the

same latitude. Inertia-gravity waves propagate outward from the impact site. The wave propagation rate is determined by the deformation radius, a key dynamical parameter that is uncertain in the troposphere. The circular coherence of the waves even after several days will facilitate detection of relatively small temperature fluctuations.

TABLE 1 Ranges of zonal temperature deviation profiles

Pressure (mbar)	10 ²⁷ erg		10 ²⁸ erg		10 ²⁹ erg		10 ³⁰ erg	
	Day 1 ΔT (K)	Day 2 ΔT (K)	Day 1 ΔT (K)	Day 2 ΔT (K)	Day 1 ΔT (K)	Day 2 ΔT (K)	Day 1 ΔT (K)	Day 2 ΔT (K)
Stratospheric energy deposition (16–69-mbar layer)								
16	0.02	0.06	0.1	0.1	0.4	0.6		
69	0.04	0.09	0.4	0.2	1.2	0.7		
287	0.02	0.08	0.2	0.1	0.7	0.5		
1,197	0.009	0.07	0.1	0.1	0.3	0.2		
5,000	0.004	0.03	0.04	0.04	0.1	0.08		
Tropospheric energy deposition (287–1,197-mbar layer)								
16	0.004	0.02	0.01	0.04	0.07	0.04	0.3	0.3
69	0.009	0.02	0.04	0.07	0.3	0.1	0.4	0.4
287	0.009	0.03	0.04	0.1	0.2	0.1	0.5	0.3
1,197	0.01	0.04	0.06	0.2	0.06	0.1	0.3	0.9
5,000	0.007	0.02	0.02	0.05	0.1	0.1	0.7	0.8

All our simulations show both a set of globally-propagating inertia-gravity waves and a longer-lived vortex at the impact site (see Fig. 2). Vortex behaviour depends on local conditions, but in our longest-run case (10²⁸ erg, stratospheric deposition, 36 d) the vortex sheared into two components that moved west-north-west and east-southeast. In all simulations inertia-gravity waves travel at $\sim 400 \text{ m s}^{-1}$ in the stratosphere and slower in the troposphere. There is an antipodal wave crossing, but no single focus of energy because of wave dispersion and Jupiter's oblateness. Because there is less mass and stronger stratification, stratospheric deposition excites stronger waves than tropospheric deposition (see Fig. 1). Table 1 contains each simulation's temperature deviation range. The waves interfere with one another behind the first wavefront, making it difficult to determine the functional form of the increase of temperature range ΔT with comet energy, although it is not faster than linear.

Observers can probe different atmospheric levels by selecting wavelengths that are sensitive to different regions. The predicted temperature deviations caused by Shoemaker–Levy 9 inertia-gravity waves bracket the detection threshold for thermal infrared imaging, which probes the stratosphere^{22–24}. Transient condensation effects similar to those seen in mountain lee waves may be visible in high-resolution, reflected-light images of the waves' horizontal passage through the ammonia clouds (upper troposphere). Whether such effects are indeed seen will depend upon local circumstances and the size of the wave. Inertia-gravity waves will be distinguishable from seismic waves¹⁹ by the former's slower propagation rate and stronger thermal signature. Observers should be careful not to confuse the remnant signatures of seismic effects close to the impact sites with those of propagating inertia-gravity waves. In all cases, techniques that take advantage of the waves' circular coherence (see Fig. 2), such as averaging in radial bins around the impact locations, will improve the signal-to-noise ratio of the wave trace. Because of the uniqueness of this event and the potential for significant improvements in dynamical modelling, observers should attempt to determine inertia-gravity wave speeds at as many pressure levels as possible. □

17. Pirraglia, J. A., Conrath, B. J., Allison, M. D. & Gierasch, P. J. *Nature* **292**, 677–679 (1981).
18. Limaye, S. S. *Icarus* **65**, 335–352 (1986).
19. Marley, M. S. *Astrophys. J. Lett.* (in the press).
20. Klemp, J. B. & Lilly, D. K. *J. Atmos. Sci.* **35**, 78–107 (1978).
21. Ahrens, T. J., Takata, T. & O'Keefe, J. D. *Bull. Am. Astr. Soc.* **25**, 1043 (1993).
22. Deming, D. et al. *Astrophys. J.* **343**, 456–467 (1989).
23. Gezari, D. Y. et al. *Nature* **342**, 777–780 (1989).
24. Orton, G. S. et al. *Science* **252**, 537–542 (1991).

ACKNOWLEDGEMENTS. We thank E. Charrette and the MIT Earth Resources Laboratory for programming assistance and use of their nCUBE 2 massively parallel computer. This work was funded by NASA.

Trapped electron acceleration by a laser-driven relativistic plasma wave

M. Everett, A. Lal, D. Gordon, C. E. Clayton, K. A. Marsh & C. Joshi

Department of Electrical Engineering,
University of California at Los Angeles,
Los Angeles, California 90024, USA

THE aim of new approaches for high-energy particle acceleration¹ is to push the acceleration rate beyond the limit ($\sim 100 \text{ MeV m}^{-1}$) imposed by radio-frequency breakdown in conventional accelerators. Relativistic plasma waves, having phase velocities very close to the speed of light, have been proposed^{2–6} as a means of accelerating charged particles, and this has recently been demonstrated^{7,8}. Here we show that the charged particles can be trapped by relativistic plasma waves—a necessary condition for obtaining the maximum amount of energy theoretically possible for such schemes. In our experiments, plasma waves are excited in a hydrogen plasma by beats induced by two collinear laser beams, the difference in whose frequencies matches the plasma frequency. Electrons with an energy of 2 MeV are injected into the excited plasma, and the energy spectrum of the exiting electrons is analysed. We detect electrons with velocities exceeding that of the plasma wave, demonstrating that some electrons are 'trapped' by the wave potential and therefore move synchronously with the plasma wave. We observe a maximum energy gain of 28 MeV, corresponding to an acceleration rate of about 2.8 GeV m^{-1} .

Relativistic plasma waves are particularly useful for accelerating particles for two reasons. First, the electric fields within these waves can be extremely high. Second, the fact that these waves can be made to move at the same speed as a highly relativistic particle means that the accelerating particles will interact with the fields for a long time, and can thus be accelerated to very high energies. In this 'plasma beat-wave accelerator' experiment, such a wave is driven by illuminating a suitable plasma with two laser beams of slightly different frequencies. The interference—or beating—of the two beams exerts a force on the plasma at the difference frequency $\Delta\omega$ and difference wavenumber Δk , thus modulating the plasma density. The electric field of the wave is just the space-charge field associated with this modulated electron density as one would find from Gauss's law.

Over the last decade, many groups have reported success in the excitation of relativistic plasma waves using the 'beat-wave' technique^{9–11}. More recently, acceleration of externally injected⁷ as well as background plasma electrons⁸ by beat-excited plasma waves has also been shown. But in order to make a useful device based on this scheme, not only must electrons be injected externally into such a wave, they must be 'trapped' by the wave potential¹². Although trapping is not a necessary condition for some energy gain, only the trapped electrons can gain the theoretical maximum amount of energy limited by dephasing which occurs when the polarity of the electric field of the plasma wave

Received 14 December 1993; accepted 3 March 1994.

1. Nakano, S. *IAU Circ. No.* 5800 (1993).
2. Carusi, A. & Yeomans, D. K. *IAU Circ. No.* 5807 (1993).
3. Chodas, P. W. & Yeomans, D. K. *Bull. Am. Astr. Soc.* **25**, 1042–1043 (1993).
4. Dowling, T. E. *Bull. Am. Astr. Soc.* **25**, 1031 (1993).
5. Scotti, J. V. & Melosh, H. J. *Nature* **365**, 733–735 (1993).
6. Weaver, H. A. et al. *Science* **263**, 787–791 (1994).
7. Sekanina, Z. *Science* **262**, 382–387 (1993).
8. Zahnle, K. & Mac Low, M.-M. *Icarus* (in the press).
9. Conrath, B. J., Gierasch, P. J. & Nath, N. *Icarus* **48**, 256–282 (1981).
10. Dowling, T. E. & Ingersoll, A. P. *J. Atmos. Sci.* **46**, 3256–3278 (1989).
11. Ingersoll, A. P. & Cuong, P. G. *J. Atmos. Sci.* **38**, 2067–2076 (1981).
12. Marcus, P. S. *Nature* **331**, 693–696 (1988).
13. Williams, G. P. & Wilson, R. J. *J. Atmos. Sci.* **45**, 207–241 (1988).
14. Melosh, H. J. & Schenk, P. *Nature* **365**, 731–733 (1993).
15. Hsu, Y.-J. G. & Arakawa, A. *Mon. Weath. Rev.* **118**, 1933–1959 (1990).
16. Dowling, T. E. *J. Atmos. Sci.* **50**, 14–22 (1993).

changes its sign. A trapped electron, by definition, moves synchronously with the wave at the point of reflection in the wave potential. At this point the trapped electron has a relativistic Lorentz factor $\gamma_{ph} = (1 - v_{ph}^2/c^2)^{-1/2}$ where v_{ph} is the phase velocity of the wave. Thereafter, as the electron continues to gain energy, it remains trapped and executes a closed orbit in the wave potential. Trapping tends to bunch the electrons in phase space resulting in a relatively narrow energy spread, which is necessary for most applications. Trapped electrons bunch in physical space because the trailing electrons see a higher field than the leading electrons and tend to catch up with the latter. This bunching then leads to a narrow energy spectrum as all the electrons subsequently interact with roughly the same accelerating field. If the initial electron energy is less than $0.511(\gamma_{ph} - 1)$ MeV (the kinetic energy of an electron moving synchronously with the wave), observation of accelerated electrons with energies greater than this indicates acceleration of trapped electrons by the wave. Here we present the first evidence of such trapped particle acceleration of externally injected electrons by a relativistically propagating plasma wave.

The experimental parameters and the apparatus has been discussed in detail elsewhere¹³. A two-frequency CO₂ laser beam (10.28 μm and 10.59 μm wavelengths, 150 ps rise time and $5 \times 10^{14} \text{ W cm}^{-2}$ peak intensity) and a co-propagating 2-MeV electron beam are both focused to the same point in a vacuum chamber filled with $\sim 155 \text{ mT}$ of hydrogen. The gas pressure is chosen so that the difference frequency between the two laser lines $\Delta\omega$ equals the plasma frequency ω_p of the laser-ionized gas. This plasma frequency corresponds to a plasma-wave wavelength of only 0.34 mm. On full ionization, the plasma wave grows rapidly in time and some of the injected electrons are trapped and accelerated. The injected electrons have a full width at half maximum (FWHM) spot size of 250 μm whereas the laser spot size is $\sim 180 \mu\text{m}$ (FWHM). The average electron current over the 1.5 ns (FWHM) macropulse is 25 mA. This macropulse has a microstructure at 9.3 GHz, with each micropulse $\sim 10 \text{ ps}$ (FWHM) in duration, giving a peak current of 270 mA. Because the macropulse is much longer in duration than the

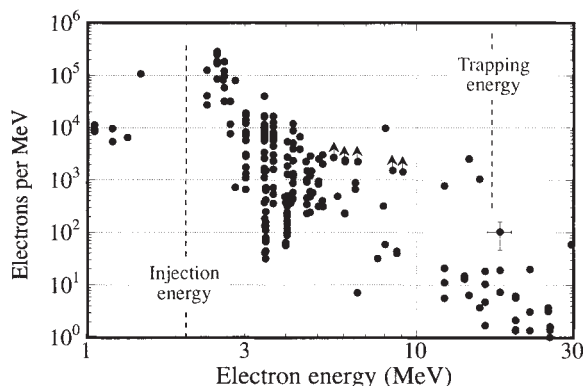


FIG. 1 The cumulative energy spectrum of electrons leaving the plasma wave within a full cone angle of 6° . To obtain this data, various charge-sensitive preamplifiers and thick-lead apertures were used to extend the dynamic range of the electron detectors to over four orders of magnitude in electron flux beginning with single-electron sensitivity. The energy dispersion across the aperture of the detector was considered in arriving at the plotted electrons per MeV. The points with arrows are lower bounds due to detector saturation. At a given setting of the electron-spectrometer magnetic field, data can only be obtained at three energies per laser shot. Therefore, a complete electron spectrum has to be accumulated over many shots. The dashed lines mark the injection energy and the trapping, where the latter refers to the energy which the electrons must have to move synchronously with the plasma wave.

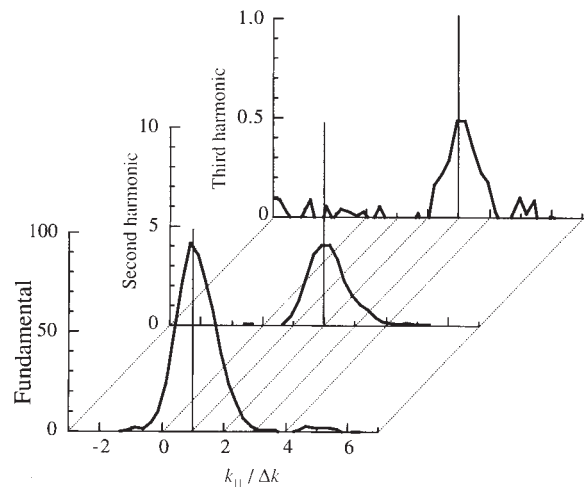


FIG. 2 Signals at $\Delta\omega$, $2\Delta\omega$ and $3\Delta\omega$ taken from the time-integrated frequency-wavenumber spectrum of the plasma wave obtained by small-angle Thomson scattering. The width of each peak is instrument-limited; the height is given in arbitrary units. Because the wavelength of the plasma wave and its radial extent are comparable, the plasma wave has a broad perpendicular wavenumber ($k_\perp$) spectrum. In the 90° scattering geometry used here, energy and momentum can only be conserved¹⁷ for photons that scatter from the component of the wave with $k_\perp = k_\parallel (=m\Delta k$ for the m th harmonic scattered order). From knowledge of the laser spot profile, we estimate the $k_\perp$ spectrum of the plasma wave and thereby quantify the relative amplitude of each wave harmonic¹⁸.

excited plasma wave, most of the injected electrons are unaffected by the plasma wave. These electrons are dumped onto low-density plastic. The accelerated electrons leave the vacuum through a 25- μm -thick Mylar window and are detected electronically by one or more silicon surface barrier detectors or by a gaseous Cherenkov cell followed by a photomultiplier tube. Electrons are also detected photographically by the tracks they leave in a cloud chamber⁷.

The two main diagnostics are an electron energy spectrometer and 'small-angle' Thomson scattering^{14,15}. A complete spectrum of electrons that are accelerated—as well as those decelerated (because they see an electric field of the wrong polarity)—by the plasma wave is shown in Fig. 1. The large scatter in the absolute number of electrons in a given energy bin is due both to the timing jitter ($\pm 50 \text{ ps}$) between the injected electron bunch and the plasma wave, and to the shot-to-shot variation in the laser pulses. The lifetime of the accelerating fields estimated from the optical measurements described below is $\sim 50 \text{ ps}$ (FWHM). Thus the points that give the maximum number of electrons at a given energy probably correspond to perfect synchronism of the injected electrons with the peak plasma wave fields. The jitter also contributes to the large energy spread observed in the spectrum. Another contributing factor to the spread is that the electrons are not pre-bunched into the ideal phases of the wave. The plasma wave period is $\sim 1 \text{ ps}$ whereas the electron bunch is about 10 ps and therefore the injected electrons uniformly occupy all phases over ~ 10 plasma oscillations. In these experiments the plasma wave has a Lorentz factor γ_{ph} of ~ 33 which would be synchronous with a 16-MeV electron. Therefore all the observed electrons with energies above 16 MeV are indeed trapped electrons and are thus moving forward in the frame of the wave.

The maximum electron energy gives an upper bound on the longitudinal electric field of the wave integrated along the electron trajectory through the wave. Because (as discussed earlier) the electric field is the space-charge field arising from plasma density perturbations of the wave, optical Thomson scattering

can be used to measure these density perturbations and obtain information about the wave electric fields independent of the energy gain measurement. Essentially, the density perturbations look like a moving volume grating which can diffract the probe beam into frequency-shifted 'orders'. The 2-ns (FWHM), 50-MW, 0.53- μm optical probe beam ($\omega_{\text{pr}}, k_{\text{pr}}$) traverses the plasma wave at a right angle to its direction of propagation. In this configuration, the probe beam has a cylindrical focus 6 mm (~ 20 plasma wavelengths) long. The scattered light is observed to be emitted at various discrete angles $\theta = k/k_{\text{pr}}$ (where $k = m\Delta k$) and frequency-shifted by $\omega_{\text{pr}} - m\omega_{\text{p}}$. Here $k_{\parallel}$ is the wavenumber of the m th harmonic of the plasma wave along the propagation direction of the electron beam and Δk is the wavenumber difference between the two laser beams. Figure 2 shows the time-integrated ω - k spectrum of the density perturbations of the wave obtained by Thomson scattering. This spectrum clearly shows discrete peaks at $(m\Delta\omega, m\Delta k)$ for $1 \leq m \leq 3$ corresponding to scattering of light from the fundamental and second and third harmonics of the relativistic plasma wave. Using nonlinear wave theory¹⁶, which gives the normalized amplitude of the m th har-

monic a_m as a function of the normalized amplitude of the fundamental a_0 as $a_m \approx a_0^m$, one can obtain an estimate of the plasma wave amplitude based on its harmonic content. After correcting for the finite- $k_{\perp}$ effects^{17,18}, we obtain an estimate for the density perturbation of 38% of the background density (based on the second-harmonic to fundamental ratio), or 40% (from the third-harmonic to fundamental ratio). Alternatively, we can obtain an amplitude estimate based on the absolute intensity of the scattered light from which we infer a 30% wave amplitude. This latter estimate agrees well with two-dimensional 'particle-in-cell' computer simulation of our exact experimental parameters. We can compare these optical measurements to the electron measurements by noting that the interaction length for our focusing conditions is expected to be ~ 1.0 cm (ref. 19). Thus, the 30-MeV electrons gained energy at a rate of 2.8 GeV m^{-1} which, from Gauss's law, corresponds to a wave amplitude of 28%, in good agreement with the independent optical diagnostic and the computer simulation. A more theoretical discussion of the acceleration and scattering of electrons observed in our experiments will be presented elsewhere¹⁹. $\square$

Received 15 November 1993; accepted 22 February 1994.

1. Sessler, A. M. *Physics Today* **41**, 26–34 (1988).
2. Dawson, J. M. *Scient. Am.* **260**, 54–61 (1989).
3. Trajima, T. & Dawson, J. M. *Phys. Rev. Lett.* **43**, 267–270 (1979).
4. Sprangle, P., Esarey, E., Ting, A. & Joyce, G. *Appl. Phys. Lett.* **53**, 2146–2148 (1988).
5. Chen, P., Dawson, J. M., Huff, R. W. & Katsouleas, T. *Phys. Rev. Lett.* **54**, 693–696 (1985).
6. Joshi, C. *et al. Nature* **311**, 525–529 (1984).
7. Clayton, C. E. *et al. Phys. Rev. Lett.* **70**, 37–40 (1993).
8. Kitagawa, Y. *et al. Phys. Rev. Lett.* **68**, 48–51 (1992).
9. Clayton, C. E., Joshi, C., Darrow, C. & Umstadter, D. *Phys. Rev. Lett.* **54**, 2343–2346 (1985).
10. Amiranoff, F. *et al. Phys. Rev. Lett.* **68**, 3710–3713 (1992).

11. Dangor, A. E., Dymoke-Bradshaw, A. K. L. & Dyson, A. E. *Physica Scripta* **T30**, 107–109 (1990).
12. Williams, R. L., Clayton, C. E., Joshi, C., Katsouleas, T. & Mori, W. B. *Lasers Particle Beams* **8**, 427–449 (1990).
13. Joshi, C. *et al. Advanced Accelerator Concepts* 379–410 (AIP Conf. Proc. No. 279, Amer. Inst. Phys., New York, 1993).
14. Sheffield, J. *Plasma Scattering of Electromagnetic Radiation* (Academic, New York, 1975).
15. Clayton, C. E., Darrow, C. & Joshi, C. *Appl. Opt.* **24**, 2823–2826 (1985).
16. Jackson, E. A. *Phys. Fluids* **3**, 831–833 (1960).
17. Martin, F., Johnston, T. W. & Ebrahim, N. *Phys. Rev. Lett.* **55**, 1651 (1985).
18. Clayton, C. E., Joshi, C., Darrow, C. & Umstadter, D. *Phys. Rev. Lett.* **55**, 1652 (1985).
19. Clayton, C. E. *et al. Phys. Plasmas* (in the press).

Emission of blue light from hydrogenated amorphous silicon carbide

W. A. Nevin, H. Yamagishi, M. Yamaguchi & Y. Tawada

Central Research Laboratories, Kaneka Corporation, 1-2-80, Yoshida-cho, Hyogo, Kobe 652, Japan

THE development of new electroluminescent materials is of current technological interest for use in flat-screen full-colour displays¹. For such applications, amorphous inorganic semiconductors appear particularly promising, in view of the ease with which uniform films with good mechanical and electronic properties can be deposited over large areas². Luminescence has been reported¹ in the red-green part of the spectrum from amorphous silicon carbide prepared from gas-phase mixtures of silane and a carbon-containing species (usually methane or ethylene). But it is not possible to achieve blue luminescence by this approach. Here we show that the use of an aromatic species—xylene—as the source of carbon during deposition results in a form of amorphous silicon carbide that exhibits strong blue luminescence. The underlying structure of this material seems to be an unusual combination of an inorganic silicon carbide lattice with a substantial 'organic' π -conjugated carbon system, the latter dominating the emission properties. Moreover, the material can be readily doped with an electron acceptor in a manner similar to organic semiconductors³, and might therefore find applications as a conductivity- or colour-based chemical sensor.

The hydrogenated amorphous silicon carbide ($\text{a-Si}_{1-x}\text{C}_x\text{:H}$) films were deposited from mixtures of xylene (C_8H_{10} , Fig. 1 inset) and silane (SiH_4). Figure 1 shows the infrared spectra of three films, fabricated using different compositions of reaction

gas mixture. Vibrational modes typical of those found in $\text{a-Si}_{1-x}\text{C}_x\text{:H}$ prepared from conventional nonaromatic carbon sources, such as methane⁴, are seen. These are characteristic of an inorganic lattice, where silicon and carbon are tetrahedrally coordinated in a covalent sp^3 -bonded system, with hydrogen acting as a network terminator in the form of CH_n and SiH_m groups⁵. As expected, the ratio of the intensities of the C–H to Si–H modes decreases with decreasing xylene gas fraction (and hence, x), in the order of Fig. 1a–c. In addition, the presence of large amounts of π -bonded carbon is indicated by the occurrence of strong absorption due to sp^2 C–H stretching and sp^2 C=C stretching modes⁶. The sharpness of these modes and the occurrence of additional sharp peaks in the region of $600\text{--}800 \text{ cm}^{-1}$, at positions similar to those of pure xylene, implies that the bulk of the sp^2 -carbon sites are in the form of relatively isolated aromatic rings having a high degree of bonding to hydrogen. This is in marked contrast to films prepared from conventional sources, where the C–H modes consist almost entirely of sp^3 -carbon sites, and although an increasing amount of sp^2 carbon is seen at contents above $x \approx 0.6$, this has an unhydrogenated, fused-ring, graphitic-type structure^{4,5}. It should be noted that our films do also contain some olefinic and/or fused aromatic structure, shown by the presence of a broad underlying absorption in the sp^2 C–H and C=C stretching regions⁶.

The optical band gap (E_{opt}) for the range of films prepared is plotted as a function of carbon content in Fig. 2. Two approximately linear regions are seen with a change to a steeper slope at $x \approx 0.75$. This is in good agreement with theory, which predicts an increase in slope when the band edges change from Si-like to SiC-like as x increases⁵. In contrast, in films prepared from conventional sources, E_{opt} generally reaches a maximum value of $<3.0 \text{ eV}$ at $x \approx 0.60$, then decreases⁵. This decrease is thought to result from a change in the band-edge character from one controlled by the σ -bonded SiC lattice to a $(\pi-\pi^*)$ orbital-type gap, related to the increasing presence of sp^2 graphitic clusters. The $(\pi-\pi^*)$ band gap is expected to decrease with increasing

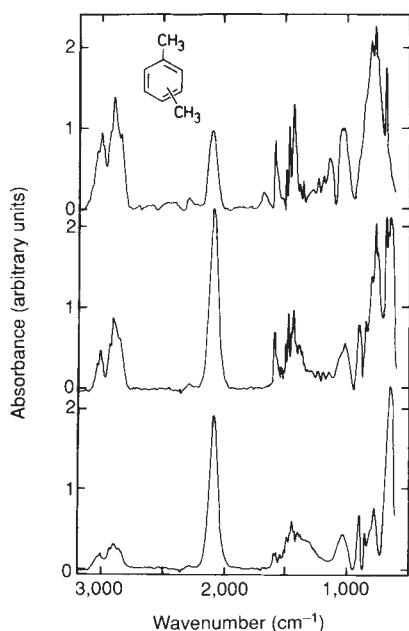


FIG. 1 Fourier-transform infrared spectra of $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ films on crystalline silicon substrates, deposited using xylene gas fractions of 0.64 (a), 0.17 (b) and 0.02 (c). The inset shows the chemical structure of the xylene molecule. Total carbon contents of these films, measured by X-ray photoelectron spectroscopy, are $x = 0.88$, 0.60 and 0.44, respectively. The films were fabricated at 200 °C by radio-frequency glow discharge decomposition of mixtures of pure xylene vapour and pure silane gas in a parallel-plate capacitively-coupled reactor operated at 13.56 MHz, under a gas pressure of 0.55 torr and with radio-frequency power of 32 mW cm^{-2} (ref. 15). The xylene gas fraction in the reaction mixture is defined as the xylene flow rate divided by the total gas flow rate. The silane flow rate was kept constant at 5 standard $\text{cm}^3 \text{min}^{-1}$. Film thicknesses were $\sim 1\text{--}2 \mu\text{m}$. Important inorganic vibrational modes lie at 2,800–3,000 cm^{-1} ($sp^3 \text{CH}_n$ [$n = 1\text{--}3$] stretching), 2,100 cm^{-1} (SiH_m [$m = 1\text{--}2$] stretching), 1,000–1,500 cm^{-1} ($sp^3 \text{CH}_n$ bending/wagging), 700–900 cm^{-1} (Si–C stretching) and 660 cm^{-1} (SiH_m wagging), while principal organic modes are seen at 3,000–3,100 cm^{-1} ($sp^2 \text{C-H}$ stretching) and 1,400–1,600 cm^{-1} ($sp^2 \text{C=C}$ stretching). The percentage of $sp^2 \text{C-H}$ is estimated (from the peak areas) to be $\sim 30\%$ of the total C–H, varying only slightly with total carbon content.

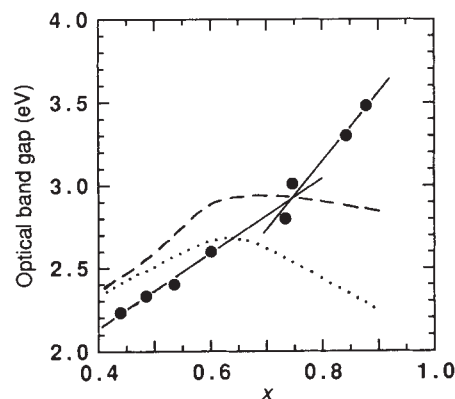
cluster size^{5,7}, so that at a carbon content of $x \approx 0.6\text{--}0.7$, the valence and conduction band edges of the π -electron system move within the sp^3 -bonded band gap. In our samples, however, the optical gap continues to increase with carbon content, despite the large amount of sp^2 sites, implying that the band edges remain controlled by the inorganic lattice throughout. The different behaviour may be due to the highly dispersed, weakly-fused nature of the aromatic network in our films, so that the energy gap (E_g) of the π -orbitals remains large, at least more than $\sim 3.5 \text{ eV}$ (note that the first electronic transition of xylene lies at 4.55 eV (ref. 8)). Using the relationship $E_g \approx 6M^{-1/2} \text{ eV}$, where M is the number of sixfold aromatic rings in a cluster^{5,7}, it is estimated that the largest aromatic clusters in our films contain less than three fused rings. We visualize a predominant structure of an inorganic $a\text{-SiC}$ matrix containing an interdispersion of hydrogenated π -delocalized aromatic units (which consist of xylene-related rings and small-sized clusters) with possibly some small degree of π -orbital overlap between these units.

In view of the substantial π -bonded carbon in the films, we investigated the possibility of doping with an electron acceptor in a similar manner to organic semiconductors³. Figure 3 shows the effect of exposure to iodine vapour on the ultraviolet-visible absorption spectrum of a film with an optical gap of 3.0 eV. A strong, broad band at 3.44 eV (360 nm) appears which increases with increasing exposure time to saturation after $\sim 50\text{--}60 \text{ min}$. At the same time, the colour of the film changes from colourless to yellow-orange. This new peak is probably the result of a

charge-transfer interaction, in which an electron transfers from an upper π -orbital of a carbon π -conjugated unit in the $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ film to an iodine molecule. Charge-transfer bands have been observed for the interaction of iodine with many aromatic molecules, such as xylene or naphthalene in solution⁸ (at 3.95 and 3.45 eV, respectively). In this type of process, a complex is formed (between the electron-acceptor and -donor) which is partially ionized in its ground state. Under excitation by light, an electronic transition occurs to an excited level, in which the complex exists in a predominantly ionized state, resulting in a typically intense absorption band.

As shown in the inset of Fig. 3, the exposure to iodine results (in conjunction with the spectroscopic changes) in an enormous increase in the electrical conductivity of the films. This effect is probably also the result of a charge-transfer interaction, with a mechanism similar to that in iodine-doped organic semiconductors^{3,9}. In a speculative mechanism, an electron in the highest occupied molecular π -orbital of a π -bonded unit is thermally excited into an accepting level in the iodine molecule, leaving a hole in the π -orbital valence level. Under the influence of an applied electric field, the electron and hole can separate, with the hole then transferring by hopping between adjacent π -conjugated units. The resulting conductivity (of the order of $10^{-4} \text{ S cm}^{-1}$) compares favourably with values of $\sim 10^{-5}$ and $10^{-3} \text{ S cm}^{-1}$ for boron (p-type) and phosphorous (n-type) doping, respectively, of conventional $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ films of $E_{\text{opt}} \approx 1.9 \text{ eV}$; it is significantly greater than the conductivity of

FIG. 2 Dependence of optical band gap (E_{opt}) on carbon content x for $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ films fabricated on Corning 7059 glass substrates at 200 °C using xylene gas fractions of 0.02–0.64. The solid lines are least-squares fits through the first six and last four data points. The dashed and dotted lines show typical data for conventional $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ prepared using methane and ethylene, respectively (from ref. 5). Values of E_{opt} were determined from the optical absorption edges using Tauc plots, according to the relationship $(\alpha h\nu)^{1/2} = B^{1/2}(h\nu - E_{\text{opt}})$, where α is the optical absorption coefficient, $h\nu$ is the photon energy and B is the Tauc coefficient. Carbon contents were determined by X-ray photoelectron spectroscopy. The refractive indices of films having $x = 0.4$, 0.6 and 0.9 are 2.3, 1.85 and 1.6, respectively, and the $B^{1/2}$ values of these films are 630, 460 and 400 $\text{cm}^{-1/2} \text{ eV}^{-1/2}$, respectively. The values at high x indicate a fairly porous, disordered structure as found in previous high-carbon-content $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ films.



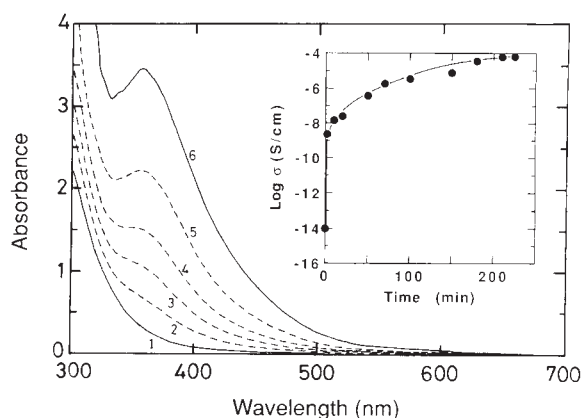


FIG. 3 Main graph, change of ultraviolet-visible absorption spectrum, measured in air, after successive 10-min exposures to iodine vapour for an $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ film with an optical band gap of 3.0 eV (Trace 1 is initial spectrum, no exposure). Inset, change of electrical conductivity (σ) for a similar film (different experiment) during exposure to iodine vapour. Both experiments were carried out at room temperature. The electrical measurements were made using coplanar evaporated gold electrodes, which were found to give ohmic contacts at the level of applied voltage used. Similar, but faster, spectroscopic and conductivity effects are observed with films of band gap 3.5 eV, with saturation occurring at ~ 40 min exposure. In addition, these kinds of changes can also be effected by exposing the films to iodine solution. The doping effect of iodine is slowly reversible in air or more rapidly by warming under vacuum, apparently due to the simple diffusion of iodine out of the sample. From the intensity of the charge-transfer peaks at saturation, taking a typical extinction coefficient of $10^4 \text{ mol}^{-1} \text{ cm}^{-1}$ for the charge-transfer band of xylenes and other aromatic molecules with iodine⁸, and assuming that all the iodine molecules are complexed, the content of iodine in the films is estimated to be of the order of 10^{21} cm^{-3} , which is of the same order as the density of aromatic rings in the films.

doped films of $E_{\text{opt}} > 2 \text{ eV}$ ($< 10^{-6} \text{ S cm}^{-1}$).

The material also exhibits an unusual excited-state behaviour. As shown in Fig. 4, films with optical band gaps of 2.6–3.5 eV give blue photoluminescence at room temperature, strongly visible to the naked eye, with a peak energy of 2.61 eV. The highest peak energy previously reported for an $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ alloy was 2.48 eV for films prepared from tetramethylsilane, which gave white luminescence and electroluminescence^{10–12}. The xylene-based films are therefore the first examples of blue-light-emitting $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ material, while the emission is also bluer than in previous amorphous carbon ($a\text{-C:H}$) films. In addition, the bandwidth is relatively narrow, about half that of previously reported high-energy $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ luminescence^{1,10–12}. Because the emission energy remains constant for optical gaps ranging from 2.6 to 3.5 eV, it is unlikely that the radiative transition involves states associated with the σ -bonded inorganic lattice, the energy of which should vary with the band gap. Instead, by analogy with the strong, room-temperature emission seen in $a\text{-C:H}$ due to sp^2 graphitic carbon clusters^{7,13}, we tentatively attribute the blue emission to the radiative recombination of optically

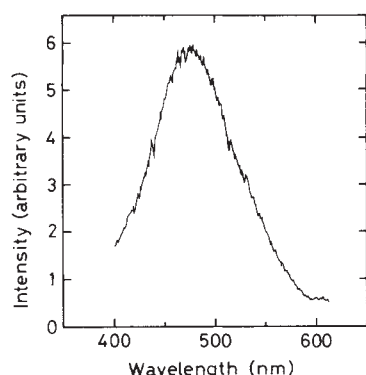


FIG. 4 Room-temperature photoluminescence (PL) spectrum of an $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ film with an optical band gap of 3.5 eV. The measurement was made using an excitation source consisting of a Xe/Hg lamp and a 3.44 eV (360 nm) narrow bandpass filter. Films with optical band gaps of 3.3, 3.0, 2.8 and 2.6 eV give similar-shaped PL spectra having a constant peak energy of 2.61 eV (475 nm), and with a 10-fold decrease in intensity on going from the 3.5 to 2.6 eV films (the emission intensity was corrected for the absorption coefficient, α , at 3.44 eV by dividing the measured value by $(1 - \exp(-\alpha d))$, where d is the film thickness). For films with a band gap below 2.6 eV ($x < 0.6$), the room-temperature PL becomes rapidly weaker and is replaced by a temperature- and composition-dependent emission peak at $\sim 1.77 \text{ eV}$ ($\sim 700 \text{ nm}$), which is strongly quenched at room temperature.

excited electron-hole pairs in localized states (such as structural defect or excimer⁸ levels) associated with the sp^2 -carbon molecular orbitals of individual π -bonded units. As the structure of the sp^2 -bonded system seems to remain fairly uniform throughout this range of samples, it is plausible that the energy levels within these localized units should also remain constant, giving rise to a constant emission energy. When the concentration of π -delocalized units falls below a critical value, at $x \approx 0.6$, the emissive mechanism seems to become dominated by a process involving sub-band-gap levels related to the sp^3 -bonded inorganic lattice, such as band tail states, as is typically found in conventional Si-rich $a\text{-Si}_{1-x}\text{C}_x\text{:H}$ alloys⁷. In this case, the emission is quenched at moderate temperatures because the main non-radiative recombination channels in the inorganic lattice are Si dangling-bond defects, so that with increased thermally-induced hopping of carriers in the band tail states, the rate of recombination at these defects increases.

To our knowledge, a material with a combination of both inorganic and organic semiconducting behaviour, such as seen here, has not been reported previously. Furthermore, the unusual properties give the material potential for a variety of technological applications. In addition to possible uses in electronic devices, the rapid uptake and release of the electron-acceptor vapour, together with the high sensitivity of the electrical and spectroscopic properties, make it promising as a conductivity- or colour-based chemical sensor¹⁴. There is also a strong possibility of using the blue luminescence in blue-light-emitting $a\text{-SiC}$ -based electroluminescent devices, the development of which has not yet been realized^{1,11,12}. □

Received 12 March 1993; accepted 1 March 1994.

- Hamakawa, Y. et al. *Optoelectron. Dev. Technol.* **4**, 281–294 (1989).
- Madan, A. & Shaw, M. P. *The Physics and Applications of Amorphous Semiconductors* (Academic, London, 1988).
- Meier, H. *Organic Semiconductors* (Verlag Chemie, Weinheim, 1974).
- Tawada, Y., Tsuge, K., Kondo, M., Okamoto, H. & Hamakawa, Y. *J. appl. Phys.* **53**, 5273–5281 (1982).
- Robertson, J. *Phil. Mag.* **B66**, 615–638 (1992).
- Dischler, B. in *Proc. Eur. Mater. Res. Soc. Symp. Vol. 17* (eds Koidl, P. & Oelhafen, P.) 189–201 (Editions de Physique, Les Ulis, Paris, 1987).
- Robertson, J. *Phil. Mag.* **B66**, 199–209 (1992).
- Birks, J. B. *Photophysics of Aromatic Molecules* (Wiley, New York, 1970).
- Simon, J. & André, J.-J. *Molecular Semiconductors* (Springer, Berlin, 1985).
- Munekata, H., Shiozaki, A. & Kukimoto, H. *J. Lumin.* **24/25**, 43–46 (1981).
- Munekata, H., Murasato, S. & Kukimoto, H. *Appl. Phys. Lett.* **37**, 536–537 (1980).
- Munekata, H. & Kukimoto, H. *Appl. Phys. Lett.* **42**, 432–434 (1983).
- Viehland, J., Lin, S., Feldman, B. J., Kilgore, K. & Jones, M. T. *Solid St. Commun.* **82**, 79–81 (1992).
- Snow, A. W. & Barger, W. R. in *Phthalocyanines: Properties and Applications Vol. 1* (eds Leznoff, C. C. & Lever, A. B. P.) Ch. 5 (VCH, New York, 1989).
- Nevin, W. A., Yamagishi, H. & Tawada, Y. *J. appl. Phys.* **72**, 4989–4991 (1992).

ACKNOWLEDGEMENTS. We are grateful to A. Nakajima for assistance with the photoluminescence measurements and to N. Fukada for the X-ray photoelectron spectroscopic data. This work was supported by the New Energy and Industrial Technology Development Organization.

Patterned optical properties in photopolymerized surface-aligned liquid-crystal films

P. J. Shannon, W. M. Gibbons & S. T. Sun

Research Center, Hercules Incorporated, Wilmington, Delaware 19808-1599, USA

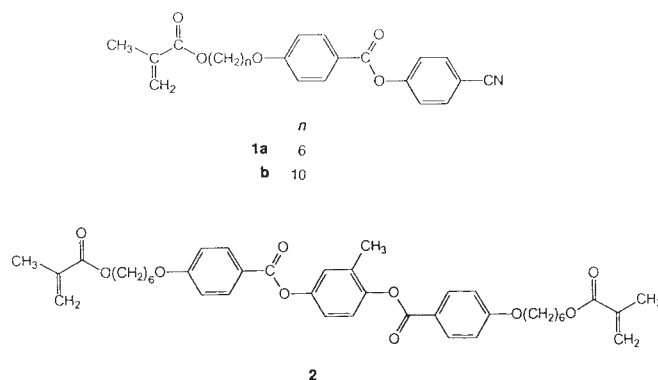
MOST practical applications of liquid crystals require control of molecular alignment at macroscopic scales¹. This is achieved most simply by confining the liquid-crystalline phase between mechanically rubbed surfaces¹. Recent developments²⁻¹¹ have shown that liquid-crystal alignment can also be controlled by optical means: for example, if azo dyes either on the alignment surface^{2,3,5,8-11} or dispersed within the liquid crystal itself^{4,6,7} are oriented by illumination with polarized light, alignment can be induced in the liquid crystal. We show here that the high-resolution alignment patterns obtainable by such techniques may be 'frozen in' by subsequent photopolymerization of the optically patterned liquid-crystalline phase. The resulting polymer films might prove valuable in the development of high-density optical storage media, three-dimensional stereo displays and other optical devices.

The basic concept of optically controlled alignment layers is illustrated in Fig. 1. A front substrate coated with a dichroic azo dye/polyimide mixture and a back substrate coated only with polyimide are mechanically rubbed and then combined to form a cell with the alignment directions of both substrates mutually parallel. Mechanical rubbing with a cloth or other fibrous material is most often used to align liquid crystals¹. The cell is then filled with a nematic liquid crystal and the rod-shaped molecules line up with the rubbing direction of the cell. The cell is then illuminated with polarized light (514.5 nm, 50 J cm⁻²) parallel to the rubbing axis, and within the absorption band of the dye. The liquid-crystal molecules adjacent to the illuminated dye-doped surface become oriented perpendicular to the direction of the electric-field polarization of the light, whereas the molecules adjacent to the undoped polyimide surface remain aligned parallel to the rubbing direction. Thus, the nematic liquid crystal assumes a twisted nematic structure within the illuminated region. It is believed² that the polarized light alters

the surface of the optically controlled alignment layer, and the liquid-crystal molecules adjacent to this layer are sensitive to this altered state and align accordingly. The illumination step can be done in the presence or absence of the nematic liquid crystal, and alignment induced by the polarized light is stable in the absence of the light field. The optically induced alignment can be subsequently erased or rewritten by simply altering the direction of the light polarization.

Aligned cholesteric¹² and nematic¹³ liquid-crystal monomer phases have been shown to undergo rapid polymerization in the presence of photoinitiators to form permanently aligned polymer films. We have used this approach to generate polymer films with a permanent high-resolution alignment pattern written into the film by optical means.

The polymerizable nematic monomer composition used here comprised a 2:2:1 (by weight) mixture of esters **1a** and **1b** (ref. 14) and the dimethacrylate crosslinker **2** (ref. 13), respectively.



Photoinitiator, inhibitor to prevent premature polymerization, and viscosity modifiers were also included. The monomers were specifically formulated to provide a stable nematic phase at room temperature; the monomer mixture had a nematic mesophase extending from 42.1 °C to below room temperature and showed no evidence of crystallization over many hours at room temperature.

The optically controlled alignment polymer was the same as described earlier, a silicone polyimide copolymer doped with a diazodiamine dye at a dye:polyimide weight ratio of 2:1 (ref. 2).

Glass substrates were spin-coated with a thin layer of polymer/dye mixture and cured. Rather than rubbing, two substrates were illuminated using an argon laser to establish the initial background alignment. A cylindrical lens was used to focus the polarized laser beam to a 1-cm line, and the substrate was scanned at a rate of 0.75 mm s⁻¹. The power was 0.9 W at 514.5 nm. One substrate was then rotated 90° with respect to the initial polarization direction, a metallized target mask (with resolution bars) attached and the scanning repeated. A cell with a 50-μm spacer was prepared with the initial polarization direction of the background illuminations mutually parallel. The cell was filled with the nematic monomer composition at 65 °C in the isotropic state, cooled in the nematic state and allowed to remain at room temperature to allow disclinations to clear. (Disclinations are optical imperfections due to the boundaries of two or more domains that differ in the orientation of the nematic director. They are evidenced by small loops or thread-like lines and are characteristic of nematic mesophases.) The cell was then irradiated with ultraviolet light from a 200 W in⁻² Hg arc lamp for 30 s to initiate polymerization.

Figure 2a, b are photomicrographs of a free-standing polymer film, independent of the original substrates, showing the imaged resolution chart as seen through parallel and crossed polarizers, respectively. The background is a homogeneous alignment (optically extinct between crossed polarizers) and the resolution bars

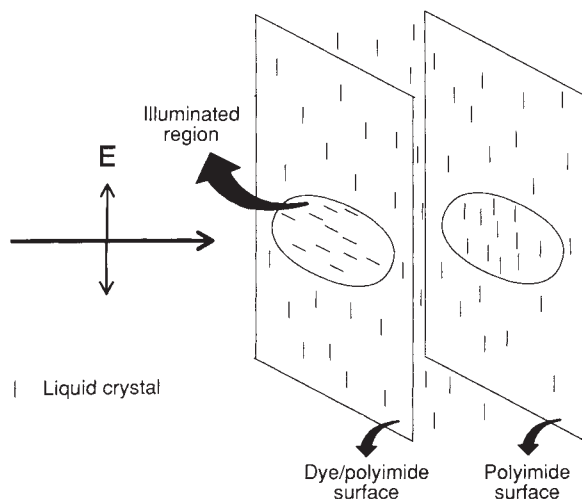


FIG. 1 The geometry of the illuminated liquid-crystal cell. The glass substrates of the cell are not shown. The rods represent the liquid-crystal orientation near the substrates before and after illumination with polarized light. 'E' indicates the direction of the electric-field polarization of the light.

have a 90° twisted alignment (transmitting between crossed polarizers). The resolution limit is $\sim 15\ \mu\text{m}$. There are several stabilized disclination defects evident in the background alignment, but they are not clear in the photographs. The quantity and size of the disclinations decrease substantially with longer annealing time, elevated temperature and lower-viscosity nematic formulations. There also are several trapped air bubbles, a result of filling the cell. The multi-aligned film is stable up to $\sim 200^\circ\text{C}$, at which point the film shatters.

The process described here amplifies an optically induced change on a surface into a bulk orientation of many molecules that can be preserved in a polymeric network. This process differs from the work of Wendorff *et al.*^{6,7}, where light-induced alignment occurs in a liquid crystal polymer controlled by azo chromophores distributed throughout the bulk of the polymer. Potential advantages include the use of liquid crystals that have low absorption at the write wavelength, and working with lower-viscosity monomers when performing the alignment. The resulting polymers have long-range alignment, and thus have optical properties similar to their low-molecular-weight liquid-crystal counterparts but without the opportunity for further mobility. The process exhibits high spatial resolution and the final crosslinked polymer network has high thermal stability.

In essence, the example we described above is a patterned polarization rotator in the form of a polymer film. This film can be used as a high-density write-once, read-only optical storage medium. The writing process defines the angle of polarization rotation at each memory location with the formation of the desired twisted local structure. In contrast to the existing binary

type of optical storage technology this patterned film of small (and potentially variable) polarization rotators increases the storage capacity many times, by the addition of multi-level, or 'grey scale', angular resolution capability¹⁵. Another application of the patterned polarization rotator is in three-dimensional stereo displays. The film can spatially separate the right and left images into two orthogonal polarizations. The stereo image will be reconstructed with a pair of eye glasses of orthogonal analysers. Furthermore, with an untwisted structure, the patterned film can be used as a binary phase device due to the birefringence of liquid crystals. The high birefringence ($\Delta n > 0.2$) achievable in liquid crystals offers the possibility of fabricating thin-film diffractive optical elements that, for example, can be used in optical interconnections¹⁶. □

Received 29 November 1993; accepted 15 February 1994.

1. Cognard, J. *Molec. Cryst. liq. Cryst.* **51**, 1–77 (1982).
2. Gibbons, W. M., Shannon, P. J., Sun, S. T. & Swetlin, B. J. *Nature* **351**, 49–50 (1991).
3. Gibbons, W. M., Shannon, P. J. & Sun, S. T. *Proc. Soc. Photo-opt. Instrum. Engrs. Symp. Liquid Crystal Materials, Devices, and Applications* Vol. 1665 184–193 (SPIE, Bellingham, Washington, 1992).
4. Sun, S. T., Gibbons, W. M. & Shannon, P. J. *Liq. Cryst.* **12**, 869–874 (1992).
5. Sun, S. T., Gibbons, W. M. & Shannon, P. J. *Proc. Soc. Photo-opt. Instrum. Engrs. Symp. Display Technologies* Vol. 1815 59–68 (SPIE, Bellingham, Washington, 1992).
6. Eich, M., Wendorff, J. H., Reck, B. & Ringsdorf, H. *Makromolek. Chem., Rapid Commun.* **8**, 59–63 (1987).
7. Anderle, K., Birenheide, R., Werner, M. J. A. & Wendorff, J. H. *Liq. Cryst.* **9**, 691–699 (1991).
8. Ichimura, K., Hayashi, Y., Akiyama, H., Ikeda, T. & Ishizuki, N. *Appl. Phys. Lett.* **63**, 449–451 (1993).
9. Ichimura, K. & Akiyama, H. *Makromolek. Chem., Rapid Commun.* **14**, 813–817 (1993).
10. Imura, Y., Kusano, J., Kobayashi, S., Aoyagi, Y. & Sugano, T. *Jap. J. appl. Phys.* **32**, L93–L96 (1993).
11. Chen, A. G. & Brady, D. J. *Appl. Phys. Lett.* **62**, 2920–2923 (1993).
12. Shannon, P. J. *Macromolecules* **17**, 1873–1876 (1984).
13. Broer, D. J., Mol, G. N. & Challa, G. *Makromolek. Chem.* **192**, 59–74 (1991).
14. Portugal, M., Ringsdorf, H. & Zentel, R. *Makromolek. Chem.* **183**, 2311–2315 (1982).
15. Marchant, A. B. *Optical Recording* 41–66 (Addison-Wesley, New York, 1990).
16. Hariharan, P. *Optical Holography: Principles, Techniques and Applications* 187–193 (Cambridge Univ. Press, Sydney, 1987).

ACKNOWLEDGEMENTS. We thank S. T. Schnelle for technical support.

Influence of atmospheric CO₂ on the decline of C4 plants during the last deglaciation

David R. Cole* & H. Curtis Monger†

*Chemical and Analytical Sciences Division,
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA
†Department of Agronomy and Horticulture,
New Mexico State University, Las Cruces, New Mexico 88003, USA

CHANGES in atmospheric carbon dioxide concentrations in the past may have caused changes in vegetation type^{1,2}, and it has been suggested² that the isotopic signature of such vegetation shifts, preserved in palaeosols³, might be used as a proxy for past CO₂ variations. But the connection between palaeosol isotopic signatures and atmospheric CO₂ concentrations has been difficult to establish, partly because of the unreliability of CO₂ proxies and partly because of the difficulty in ruling out other potential causes of vegetation changes, such as climate⁴. Here we present palaeosol carbon isotope ratios that reveal a shift from C4-dominated grasses to C3-dominated shrubs about 7–9 kyr ago on an alluvial fan system in the Chihuahuan desert, New Mexico. This coincides with a rapid increase in atmospheric CO₂ concentration recorded in Antarctic ice cores^{5–8} and increased aridity recorded by geomorphic reconstructions^{9–11} and packrat remains¹². Palaeosol oxygen isotope ratios, which depend on temperature and moisture, were relatively constant during the vegetation shift, suggesting that the CO₂ change, rather than climate, was the dominant cause. We conclude that the carbon isotope ratios of ancient soils can indeed be used as a proxy for past CO₂ changes.

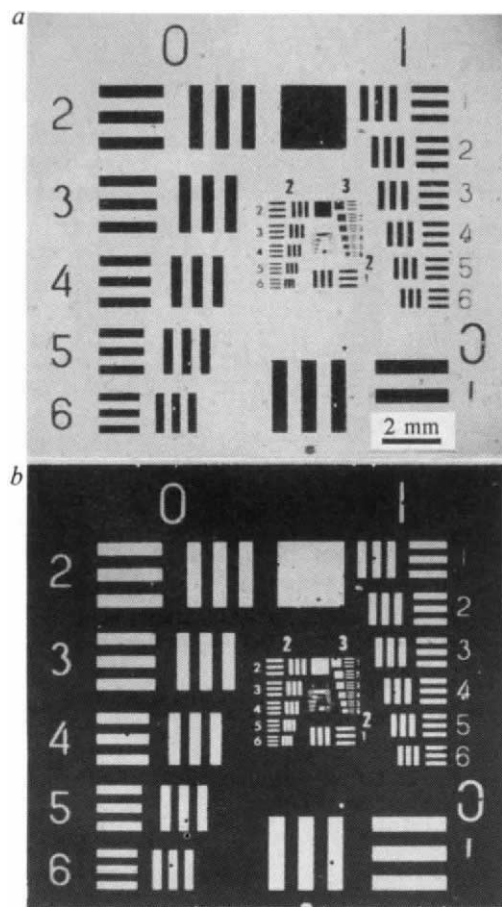


FIG. 2 Photomicrograph between parallel (a) and crossed (b) polarizers of a free-standing polymer film with multiple orientations. The background is a homogeneous alignment and the resolution bars have a 90° twisted alignment.

Radiometric dating and isotopic measurements of carbonates and organic matter from well-preserved buried soils were carried out at sites in south-central New Mexico. Four soil profiles, exposed by excavations and one arroyo-cut, are located on the southeastern piedmont-fan area of the Organ Mountains, which lie in the northern end of the Chihuahuan desert. The fan surfaces were identified and correlated to geomorphic surfaces in the neighbouring Las Cruces area by their geomorphic position¹³ and soil profile development, mainly carbonate morphology^{9,14} (Fig. 1). Based on stratigraphic criteria established by Gile *et al.*⁹ and Ruhe¹⁵, four generations of fans are recognized: Jornada I, Jornada II, Isaacks' Ranch and Organ (Fig. 1).

A wide variety of dating techniques, including ¹⁴C in charcoal and inorganic carbon, K-Ar basaltic flows and uranium-trend dating of carbonates, were used to constrain the ages of the pedogenic carbonates that accumulated in alluvial soils of this region^{9,13,16-18}. Each fan received carbonate up to the time it became buried by the next alluvial fan deposit. Once a fan became buried, the pedogenic carbonate was isolated and protected from further additions of carbonate. Pedogenic carbonates in the buried soils have undergone little or no contamination from the parent material, which is composed exclusively of silicate detritus from the Organ Mountain batholith and associated volcanics, and show no evidence of recrystallization that could lead to an isotopic resetting of the calcites. Additionally, there is an ~15‰ enrichment in $\delta^{13}\text{C}$ of soil carbonates compared to organic matter (Fig. 2), consistent with the diffusion model of Cerling³, indicating the carbonate is co-genetic with the organics and unaltered by diagenesis.

The distribution of carbon isotope values in the pedogenic calcites (Fig. 2) clearly demonstrate a depletion of $\delta^{13}\text{C}$ in Organ alluvium. Pre-Organ palaeosols, ~9,070 ¹⁴C yr BP and older, exhibit a range of -1 to -3.5‰, with an average of -2.2‰. In contrast, Organ soils have a mean $\delta^{13}\text{C}$ value of -7.8‰ for a range of -6 to -10‰. The excellent correlation observed between the trends of enrichment and depletion in $\delta^{13}\text{C}$ values of carbonates and associated organic matter is consistent with similar comparisons described by Cerling and Quade¹⁹ for modern soils and palaeosols from the midwestern and western United

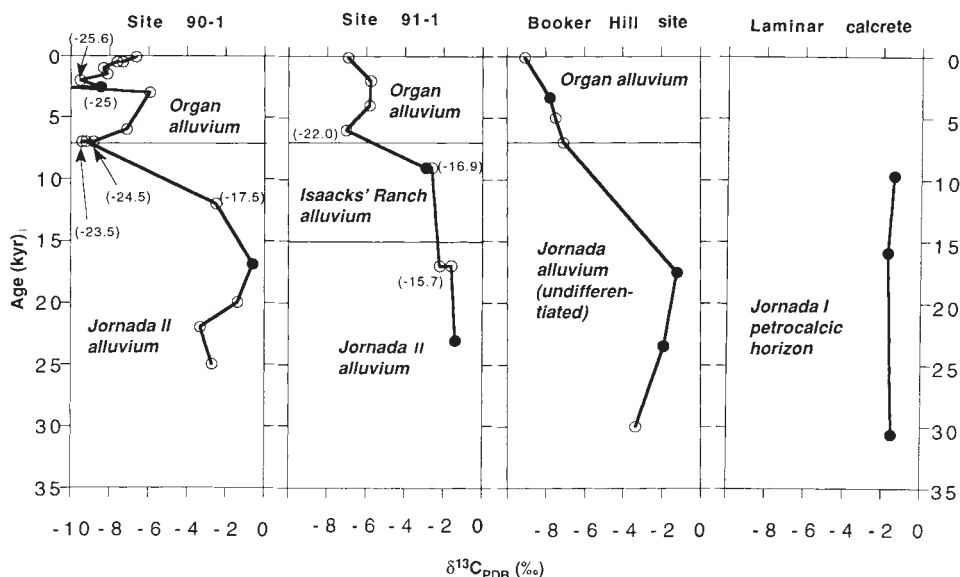
Time (kyr)	Geomorphic surfaces	Carbonate morphology	Estimated age	Dating techniques and age
10	Holocene	Organ	stage I	(9) ¹⁴ C <7kyr
		Isaacks' Ranch	stage II	(9) ¹⁴ C (18) 8-11 kyr
125	Late Pleistocene	Jornada II	stage III, IV	(17) uranium-trend 130±25 kyr
				K-Ar(max) ~180 kyr
Middle Pleistocene		Jornada I	stage III, IV	(9) uranium-trend 290±20kyr

FIG. 1 Carbonate morphology and estimated ages of geomorphic surfaces and their soils for the Las Cruces, New Mexico study area, in units of kyr before present. Petrogenic calcite progresses with age from stage I filaments and coatings through stage II and III nodules and internodular fillings to a stage IV petrocalcic horizon that can be subdivided into laminar and plugged zones¹⁸. Progressively more complex carbonate horizons occur in soils of increasingly older geomorphic surfaces. References are shown in parentheses.

States. The proposed increase^{8,20} of ~0.3–1‰ in $\delta^{13}\text{C}$ of atmospheric CO₂ from glacial to interglacial time is clearly in the opposite direction to the isotopic depletions we observe with time. This indicates that the influx of CO₂ into the soil profile was apparently very minor, and had little or no influence on the $\delta^{13}\text{C}$ trajectories of the soil carbonate.

We interpret the shift to lighter $\delta^{13}\text{C}$ values to be the result of an increase in the proportion of C3 desert shrub vegetation caused by a decline in C4 grass density, which would account for the increased erosion and onset of Organ alluviation⁹. Carbonates formed in soils dominated by soil-respired CO₂ derived exclusively from C3 plants have $\delta^{13}\text{C}$ values of between -12 and

FIG. 2 $\delta^{13}\text{C}$ of pedogenic carbonates and organic carbon plotted against age in four study sites from south-central New Mexico. (Age is derived by linear interpolation between radio-carbon dates.) Organic carbon (values given in parentheses) was isolated from the soils and analysed in a manner similar to that described by Quade *et al.*³⁴. The sites occur at an elevation of 1,275±25 m. All radio-carbon dates are indicated by solid circles and are of inorganic carbon, except for the Organ alluvium samples at 90-1 and Booker Hill sites which are of charcoal. The deposition of Organ alluvium in the study area is inferred to have started ~7 kyr BP based on stratigraphic correlation to radio-carbon-dated Organ alluvium in the neighbouring Las Cruces area⁹. The Isaacks' Ranch deposit is laterally discontinuous and only occurs at site 91-1. Samples of the 'Laminar calcrete' site were taken from accretionary layers of pedogenic carbonate in a petrocalcic horizon formed at a depth of 40 cm. Fault displacement of the 'Laminar calcrete' soil indicates that tectonic uplift has been less than a few metres during the past ~30 kyr. With the exception of the uppermost two samples from each of the soil profiles, all samples were of pedogenic carbonates that precipitated at depths >20 cm. Replicate



analyses of the $\delta^{13}\text{C}$ in carbonates were reproducible to 0.1‰; the organics have an error of ± 0.4 ‰. Duplicate samples taken 1 m apart in the same horizon varied by 0.6‰ or less for carbonates from four horizons: two in 90-1 and two in 91-1. ($\delta^{13}\text{C}_{\text{PDB}} = [({}^{13}\text{C}/{}^{12}\text{C})_{\text{sample}} / ({}^{13}\text{C}/{}^{12}\text{C})_{\text{PDB}} - 1]$ per mil.)

–10‰, with associated organic carbon ranging from –28 to –24‰ (refs 3, 21). Conversely, the $\delta^{13}\text{C}$ compositions of soil carbonate in equilibrium with CO_2 respired from C4 plants range from –3 to +1‰, with associated organic carbon ranging from –18 to –14‰. Using the approach outlined by Cerling^{3,22}, we can estimate that C4 plants comprised ~70–90% of the biomass before ~9,100 yr BP, but only 20–40% after this. Modern vegetation in the study area is dominated by C3 (60–80%) desert shrub vegetation, including mesquite and creosote-bush, with a subordinate contribution by C4 grasses (20–40%). The $\delta^{13}\text{C}$ values of between –7 and –9‰ observed in the youngest soil carbonates (~100 yr BP) are consistent with the biomass distribution occupying today's piedmont-fan landscape.

Because of an incomplete record, we are not able to constrain the exact timing of the shift from C4- to C3-dominated vegetation in the study area. But we do know that it occurred no earlier than ~9,100 yr BP (based on carbonate ^{14}C ages), and no later than ~7,000 yr BP (based on buried charcoal ^{14}C ages; this is the time that marks the boundary between younger Organ alluvium and older Jornada II or Isaacks' Ranch⁹). Based on plant macrofossils from packrat middens in the Chihuahuan desert, Van Devender²³ attributes the vegetative change at the end of early Holocene period to the decline of winter precipitation and a shift to hotter summers with monsoonal rainfall. Pluvial Lake Cochise in southeastern Arizona records a decline of effective moisture in the southwestern United States after ~8,900 yr BP (ref. 11), which agrees with climate modelling²⁴, palaeovegetational reconstruction²⁵ and soil-geomorphic relations²⁶.

A number of studies have demonstrated a good correlation between the carbon and oxygen isotope compositions of soil carbonates^{27–29}. For example lower temperatures cause a depletion of ^{18}O relative to ^{16}O in meteoric waters and also favour the C3 photosynthetic pathway, which causes ^{13}C depletion in plants. Before ~9,000 yr BP, there is a reasonably good correlation between the enrichment and depletion in both ^{18}O and ^{13}C in the pedogenic carbonates (Fig. 3a, b). It is clear that following 9,000 yr BP, the ^{13}C and ^{18}O trends become discordant: where the $\delta^{13}\text{C}$ values become very light while there is only a slight enrichment in $\delta^{18}\text{O}$ values. The $\delta^{18}\text{O}$ trends (Fig. 3a) are similar to the δD patterns of enrichment with time reported by Thompson³⁰ and Long *et al.*³¹ for waters equilibrated with plant cellulose sampled in a number of packrat middens from the Great Basin and the Colorado plateau. Using this information, we can estimate the expected $\delta^{13}\text{C}$ of pedogenic carbonates by knowing that $\delta^{13}\text{C}$ values commonly average ~3‰ higher than their ^{18}O counterpart. The estimated curve (Fig. 3b) shows a good fit to the trends determined for $\delta^{13}\text{C}$ before 9,000 yr BP, but a markedly different trajectory for pedogenic carbonates younger than this—that is, much heavier $\delta^{13}\text{C}$ values than measured (–2 to 0‰ compared to our measured values of –10 to –6‰). The absence of a correlation between $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values suggest that a variable other than climate alone was responsible for the vegetation shift.

Cerling *et al.*² invoked an argument of declining atmospheric CO_2 content to explain the apparent rapid expansion (over ~2-Myr period) of C4 ecosystems as indicated by a significant $\delta^{13}\text{C}$ increase in Late-Miocene palaeosols. But the timing of this global C4- CO_2 relationship (~7.4 Myr ago) has been called into question by Morgan *et al.*⁴ based on carbon isotope data from Pakistan and Kenya. Their $\delta^{13}\text{C}$ data, taken from enamel apatite of assorted herbivores spanning the time ~13–2 Myr, suggest that C4 grasses evolved gradually in favourable ecosystems in response to a number of environmental and physiological factors affecting photosynthetic production. Compared to studies of older systems, our work has two distinct advantages—namely, there is direct evidence (from the ice core) that a CO_2 change occurred, and we can constrain fairly accurately the timing of the vegetation shift.

What triggered this dramatic shift from C4 to C3 vegetation on the piedmont-fan area starting around 9,000 yr BP? Photores-

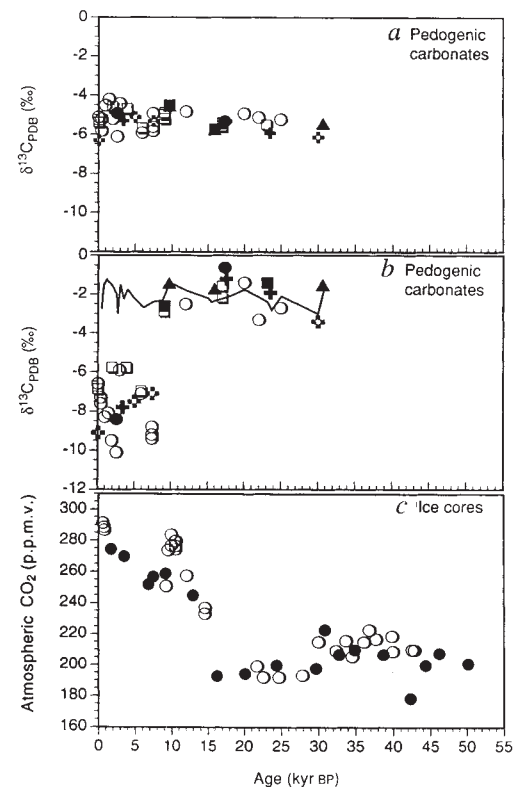


FIG. 3 Comparison of $\delta^{18}\text{O}_{\text{PDB}}$ (a) and $\delta^{13}\text{C}_{\text{PDB}}$ (b) in pedogenic carbonates and atmospheric CO_2 concentration in ice cores (c) for the period covering approximately the past 30 kyr. Symbols represent the following sites: circle (90-1); square (91-1); triangle (Laminar calcrete); cross (Booker Hill). Solid symbols represent radiocarbon-dated samples. The ages of the other samples were interpolated based on their stratigraphic position in relation to radiocarbon-dated samples. The curve drawn through the $\delta^{13}\text{C}$ data (b) is based on the assumption that the $\delta^{18}\text{O}$ data are a good proxy for relative changes in temperature (see text) and the fact that $\delta^{13}\text{C}$ values typically correlate with $\delta^{18}\text{O}$, but with a separation in magnitude of ~3‰. In (c), the solid circles represent data from the Vostok ice core samples⁶. Open circles represent data from the Byrd ice core⁸. Analyses of $\delta^{18}\text{O}$ from the replicate pedogenic samples were reproducible to $\pm 0.2\text{‰}$.

piration, which opposes photosynthesis, can become significant if atmospheric CO_2 levels drop below a few hundred p.p.m.v.¹⁹. But C4 plants use a 'pump' to transport CO_2 to the photosynthesizing sites, thus countering the threat of photorespiration more effectively than C3 plants at low atmospheric CO_2 concentrations. We propose that the increase in global CO_2 concentration from ~180 p.p.m.v. during the last glacial maximum (~18,000 yr BP) to ~275 p.p.m.v. at ~10,000 yr BP, as recorded in various ice cores^{5–8}, contributed directly to the shift from C4 to C3 vegetation. A comparison (Fig. 3b and c) of all the $\delta^{13}\text{C}$ values and representative ice-core CO_2 concentrations with age demonstrates a very good correlation between an increase of ~100 p.p.m.v. in CO_2 and a decrease in $\delta^{13}\text{C}$ starting at roughly 10,000 yr BP. This conclusion is supported by the recent experimental work of Polley *et al.*³², who grew a number of C3 and C4 plants under controlled CO_2 , temperature and moisture conditions. They observed that leaf water-use efficiency and above-ground biomass of C3 species increased linearly and nearly proportionally with a doubling of the CO_2 concentration. The results of this palaeosol study are consistent with studies of modern ecosystems³³ that indicate higher CO_2 levels favour C3 over C4 plants. This is a more satisfactory explanation of the replace-

ment of open C4 grasslands and savannahs by C3 shrublands than invoking overgrazing, suppression of fire and climate change. Moreover, the results of our study support the fact that carbon isotopic signals recorded in well-preserved, unaltered pedogenic carbonates can be used as indicators of global CO₂ change. □

Received 29 March 1993; accepted 17 February 1994.

1. Ehleringer, J. R., Sage, R. F., Flanagan, L. B. & Pearcy, R. W. *Trends Ecol. Evol.* **6**, 95–99 (1991).
2. Cerling, T. E., Wang, Y. & Quade, J. *Nature* **361**, 344–345 (1993).
3. Cerling, T. E. *Earth planet. Sci. Lett.* **71**, 229–240 (1984).
4. Morgan, M. E., Kingston, J. D. & Marino, B. D. *Nature* **367**, 162–165 (1994).
5. Stauffer, B., Neff, A., Oeschger, H. & Schwander, J. in *Geophys. Monogr.* **33** (eds Langway, C. C., Oeschger, H. and Dansgaard, W.) 85–90 (Am. Geophys. Union, Washington, DC, 1985).
6. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
7. Neff, A., Oeschger, H., Stauffer, B., Stauffer, B. *Nature* **331**, 609–611 (1988).
8. Leuenberger, M., Siegenthaler, U. & Langway, C. C. *Nature* **357**, 488–490 (1992).
9. Gile, L. H., Hawley, J. W. & Grossman, R. B. *Memoir*, No. 39 (New Mexico Bureau of Mines and Mineral Resources, Socorro, New Mexico, 1981).
10. Holliday, V. T. *Quat. Res.* **31**, 74–82 (1989).
11. Waters, M. R. *Quat. Res.* **32**, 1–11 (1989).
12. Spaulding, W. G. *Quat. Res.* **35**, 427–435 (1991).
13. Seager, W. R. *Memoir*, No. 36 (New Mexico Bureau of Mines and Mineral Resources, Socorro, New Mexico, 1981).
14. Gile, L. H., Peterson, F. F. & Grossman, R. B. *Soil Science* **101**, 347–360 (1966).
15. Ruhe, R. V. *Ann. Ass. Amer. Geogr.* **54**, 147–159 (1964).
16. Machette, M. N. in *Soils and Quaternary Geology of the SW United States* (ed. Weide, D. L.) 1–21 (Spec. Pap. No. 203, Geol. Soc. Am., Boulder, Colorado, 1985).
17. Gile, L. H. *Soil Sci. Soc. Am. J.* **51**, 752–760 (1987).
18. Gile, L. H. & Grossman, R. B. *Desert Soil Monograph* (Doc. No. PB80–13504, Natn. Inf. Serv., Springfield, Virginia, 1979).

19. Cerling, T. & Quade, J. in *Climate Change in Continental Isotopic Records* (eds Swart, P. K., Lohmann, K. C., McKenzie, J. & Savin, S.) 217–231 (Geophys. Monogr. No. 78, Am. geophys. Un., 1993).
20. Marino, B. D., McElroy, M. B., Salawitch, R. J. & Spaulding, W. G. *Nature* **357**, 461–466 (1992).
21. Cerling, T. E., Quade, J., Wang, Y. & Bowman, J. R. *Nature* **341**, 138–139 (1989).
22. Cerling, T. E. *Amer. J. Sci.* **291**, 377–400 (1991).
23. Van Devender, T. R. in *Packrat Middens: The Last 40,000 Years of Biotic Change* (eds Betancourt, J. L., Van Devender, T. R. & Martin, P. S.) 104–133 (Univ. Ariz. Press, Tucson, 1990).
24. COHMAP Members *Science* **241**, 1043–1052 (1988).
25. Spaulding, W. G. & Graumlich, L. J. *Nature* **320**, 441–444 (1986).
26. Hawley, J. W. in *Guidebook of the Las Cruces Country* (eds Seager, W. R. et al.) 139–150 (New Mexico Geol. Soc., Socorro, 1975).
27. Cerling, T. & Hay, R. L. *Quat. Res.* **25**, 63–78 (1986).
28. Wang, Y. et al., in *Climate Change in Continental Isotopic Records* (eds Swart, P. K., Lohmann, K. C., McKenzie, J. & Savin, S.) 217–231 (Geophys. Monogr. No. 78, Am. geophys. Un., 1993).
29. Mack, G. H., Cole, D. R., James, W. C., Giordano, T. H. & Salyards, S. *Am. J. Sci.* (in the press).
30. Thompson, R. S. in *Packrat Middens: The Last 40,000 Years of Biotic Change* (eds Betancourt, J. L., Van Devender, T. R. & Martin, P. S.) 200–239 (Univ. Ariz. Press, Tucson, 1990).
31. Long, A., Warneke, L. A., Betancourt, J. L. & Thompson, R. S. in *Packrat Middens: The Last 40,000 Years of Biotic Change* (eds Betancourt, J. L., Van Devender, T. R. & Martin, P. S.) 380–396 (Univ. Ariz. Press, Tucson, 1990).
32. Polley, H. W., Johnson, H. B., Marino, B. D. & Mayeux, H. S. *Nature* **361**, 61–64 (1993).
33. Mayeux, H. S., Johnson, H. B. & Polley, H. W. in *Noxious Range Weeds* (eds James, L. F., Evans, J. O., Ralphs, M. H. & Child, R. D.) 62–74 (Westview, Boulder, 1991).
34. Quade, J., Cerling, T. E. & Bowman, J. R. *Geol. Soc. Am. Bull.* **101**, 464–475 (1989).

ACKNOWLEDGEMENTS. Pollen studies by J. Gish focused our attention on the possibility of a biomass shift, and we acknowledge her interest and help in this work. Comments by J. Horita and D. J. Wesolowski are also acknowledged. Reviews by M. O'Leary and J. Hayes contributed significantly to this paper. This work was supported by the Geosciences Research Program, Office of Basic Energy Sciences, US Department of Energy (D.R.C.) and the US Army (H.C.M.).

Role of transition-metal catalysis in the formation of natural gas

Frank D. Mango*, J. W. Hightower† & Alan T. James

Departments of *Geology and Geophysics, and †Chemical Engineering, Rice University, PO Box 1892, Houston, Texas 77251, USA
Exxon Production Research Company, PO Box 2189, Houston, Texas 252–2189, USA

THE idea that natural gas is formed by thermal decomposition of sedimentary organic matter^{1,2} enjoys almost universal acceptance^{3–6}. But pyrolysis experiments on organic matter^{7–13} have failed to reproduce the composition of natural gas (typically 90% methane). It has recently been suggested¹⁴ that natural gas may instead be generated catalytically: transition metals are often found in carbonaceous sedimentary rocks, and might promote the reaction between hydrogen and *n*-alkenes (which are themselves formed during thermal decomposition of kerogen) to give light hydrocarbons and natural gas. We report here experimental results that support this hypothesis. In particular, we find that *n*-alkenes, hydrogen and a carbonaceous sedimentary rock containing moderately high concentrations of transition metals react under mild conditions (~200 °C) to generate a light-hydrocarbon product indistinguishable from natural gas in both molecular and carbon isotope composition. Our results demonstrate that the reaction is indeed catalytic, and could alter the way in which we view the generation and distribution of oil and gas in the Earth.

The sources of natural gas include biogenic gas, formed in the early stages of diagenesis, and so-called thermogenic gas, presumably generated by the thermal decomposition of organic matter^{1,2}. Here we consider thermogenic gas generated from the organic matter (kerogen and hydrocarbons) in fine-grained sedimentary rocks, and by the thermal decomposition of petroleum.

Methane dominates all forms of natural gas, varying from ~50 to 100 wt% of the C₁–C₄ fraction (Fig. 1). Any credible hypothesis on the origin of natural gas, therefore, must explain why methane is consistently the major hydrocarbon.

Laboratory pyrolysis of organic matter, however, does not give methane in the concentrations seen in natural gas. In these experiments, methane is 30–50 wt% of the C₁–C₄ product in kerogen decomposition^{7–9} and 10–40 wt% in petroleum decomposition^{10,11}. Thermal decomposition of *n*-alkanes also yields low concentrations of methane, *n*-hexadecane giving 6 wt% methane in C₁–C₃ at 600 °C, 18 wt% at 700 °C (ref. 12) and *n*-heptane yielding 13 wt% at 580 °C (ref. 13). Higher yields of methane are achieved at the higher temperatures by continued cracking of the C₂–C₄ hydrocarbons¹⁰, but we question the geological relevance of these experiments because of the extraordinary stability of these hydrocarbons. Propane, for example, has a calculated half-life of ~8 × 10⁸ yr at 200 °C¹⁵ and the decomposition products (equal parts of hydrogen, methane, ethylene and propylene) are rarely seen in natural gas.

It has been suggested that natural gas may be formed catalytically through the reaction of hydrogen and *n*-alkenes mediated by the transition metals in carbonaceous sedimentary rocks (Fig. 2)¹⁴. Oil generation from the thermal decomposition of kerogen^{16–18} may produce α -olefins which readily react with metal hydrides (M–H in Fig. 2) giving metal alkyls¹⁴, the key intermediates in Fig. 2. Hydrogen may also be a product of kerogen decomposition^{7,19}, or it may come from the reaction of water with metal-oxide mineral assemblages²⁰.

Our objective is to test the hypothesis that natural gas may be formed catalytically, promoted by transition metals through the condensation of hydrogen and α -alkenes. The intention is to employ transition metals in their natural environment, carbonaceous sedimentary rocks (source rocks), and to carry out the reactions at realistic geological temperatures. The strategy is to first exhaust a source rock of any intrinsic gas-generating capacity by first heating it in flowing hydrogen at high temperature, and then, at lower temperatures where gas generation no longer occurs, to test for catalytic activity by adding hydrogen and α -alkene. To demonstrate catalysis, we must show that any

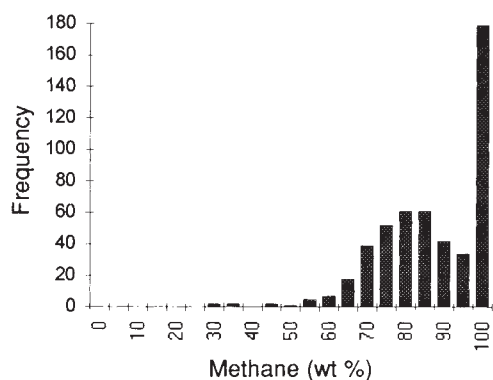


FIG. 1 Histogram of the methane concentrations (wt% in C_1-C_4) in 500 samples of natural gas randomly selected from ref. 26. Biogenic gas, which approaches 100% methane in composition, contributes to the dominant peak at 100% methane.

gas generated is coming from the added substrates (hydrogen and α -alkene) and not from the kerogen in the rock. In other words, under the time-temperature conditions in which light hydrocarbons may be generated, the kerogen in the source rock must remain inert, making no material contribution to the hydrocarbons formed. Thus, the experiments reported here should not be confused with artificial maturation experiments where kerogen is the source of carbon in the product. Our focus is on the existence, or non-existence, of catalysis in the generation of light hydrocarbons and the catalytic selectivity to methane in that process.

At 190 °C (for several days), *n*-octadecene-1, hydrogen, and source rock (preheated at 350 °C, 24 h) yielded a C_1-C_4 product that was 80 wt% methane (90 mol%). The methane was generated at a rate of the order of 10^{-7} g d $^{-1}$ g-rock $^{-1}$ and the rate remained constant over the duration of our experiments (1 yr). To verify genuine catalysis, reactions were carried out under identical conditions (190 °C, several days) with alkene + hydrogen (no rock), hydrogen + rock (no alkene), and alkene + rock (no hydrogen); in each case only traces of light hydrocarbons were detectable. The addition of moderate amounts of water (~2.4 wt% rock) appeared to increase both catalytic activity and selectivity to methane (87 wt% in C_1-C_4), and enhanced the C_{5+} product. Higher amounts of water sharply lowered the selectivity to methane with little effect on overall activity. The methane generated at 220 °C was isotopically lighter ($\delta^{13}C = -47.2\%$) than *n*-octadecene-1 ($\delta^{13}C = -30.8\%$). The -16.4‰ fractionation is within the range typically seen between so-called thermogenic gas and the *n*-alkanes in the associated petroleum^{21,22}. A commercial catalytic cracking catalyst ($SiO_2Al_2O_3$) proved active on *n*-octadecene-1 at these temperatures, but the C_1-C_4 product, 1 wt% methane and 69 wt% isobutane, was typical of acid catalysis and unlike that generated using the source rock. Figure 3 shows the gas product promoted by the source rock (Fig. 3a) compared to the products for the

same reaction with acid catalysis ($SiO_2Al_2O_3$) at 190 °C (Fig. 3b) and thermal cracking of *n*-octadecene-1 at 500 °C (Fig. 3c).

As the source rock makes no material contribution to the product but is essential for the reaction, this reaction must be viewed as catalytic. In addition, at constant substrate concentrations (that is, 1 atm hydrogen and excess α -alkene), the rate of gas generation remains constant over time (1 yr) consistent with a catalytic reaction and inconsistent with thermal cracking. In artificial maturation experiments (350 °C, 175 d), for example, where thermal cracking of kerogen occurs, the rate of gas generation remains proportional to kerogen concentration and therefore falls sharply with reaction time²³.

The source rock clearly catalyses the generation of gas through the reaction of hydrogen and α -alkenes, but are these substrates present in source rocks during oil and gas generation? The results of artificial maturation experiments, which generate liquid products remarkably like natural oils^{16, 19,23,24}, would suggest yes. The laboratory maturation of source rocks under open conditions (unsealed reactors) generates both hydrogen (10 to 70 vol% of the C_1-C_4 gas product)^{7,19} and α -alkenes^{16, 18,25} as major products. Moreover, the hydrogenation of alkenes has been proposed as the critical step in the formation of alkanes under closed conditions (sealed reactors)¹⁶. Irrespective of the source, hydrogen is a persistent component of petroleum and natural gas habitats. A conservative estimate (based on US Bureau of Mines natural-gas data) of the average concentration of hydrogen in natural gas is ~700 p.p.m. (ref. 26), and hydrogen concentrations approaching 50% have been reported in certain gas fields²⁷.

Are the experimental conditions realistic? The temperature range (190–220 °C) is clearly within the range suggested for gas generation in sedimentary basins (150–220 °C)²⁸ and significantly below the temperatures employed in artificial maturation experiments^{6,7,16–19,23–25, 28,29}. Moreover, measured reservoir temperatures up to 225 °C are reported for oil habitats³⁰ and vitrinite reflectance data from a number of deep oil and gas deposits

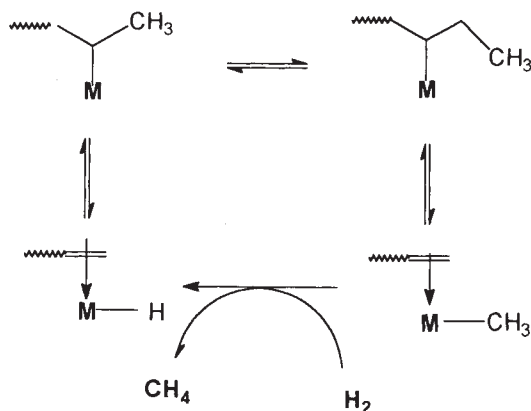


FIG. 2 The catalytic scheme proposed for the hydrogenolysis of a long-chain *n*-alkene to methane¹⁴. The wavy line denotes a long carbon chain of any length and the wavy line truncated by a double bond represents an α -alkene. The arrow from the double bond to the metal (M) indicates a weak coordinate bond. The metal maintains a σ -bond (indicated by a solid line) in each of the four intermediates shown: to H (lower left), the number two carbon of the alkyl chain (upper left), the number three carbon of the alkyl chain (upper right), and to CH_3 (lower right). The metal is free to move up or down the carbon chain by sequential addition-elimination steps. Thus, the scheme applies to the generation of ethane and higher alkanes by moving M to the number four and higher carbon atoms (for example, an ethyl group is substituted for the methyl group (lower right) when M is σ -bonded to the number four carbon of the chain (upper right)).

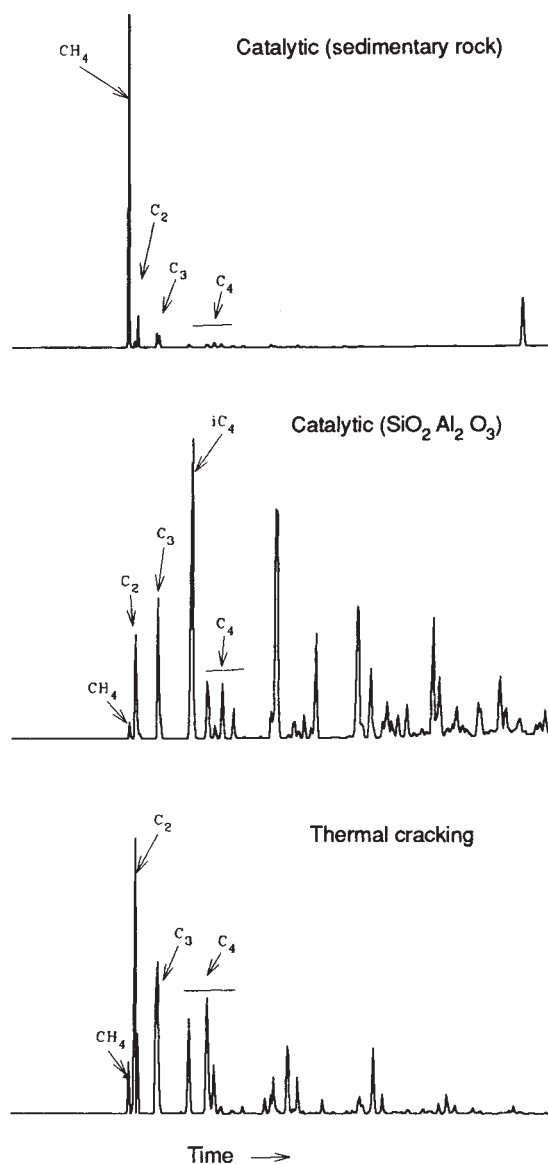


FIG. 3 a, b, c, Gas chromatograms of the products from three reactions. a, Catalytic (sedimentary rock). The gas product from *n*-octadecene-1 + hydrogen + carbonaceous sedimentary rock, heated at 190 °C for 41 d. The carbonaceous rock (Monterey Formation, California, Eocene) is a low maturity ($R_o = 0.33\%$) organic rich (14 wt% total organic carbon, 7.5 wt% sulphur, 350 p.p.m. nickel and 560 p.p.m. vanadium; Rock-Eval analysis, $S_1 = 2.48$, $S_2 = 70$; HI = 500) source rock obtained from the Cooperative Monterey Organic Geochemistry Study (KG-24)³³. b, Catalytic ($\text{SiO}_2\text{Al}_2\text{O}_3$). A catalytic cracking catalyst (Houdry, M-46) was heated under flowing hydrogen at 250 °C for 24 h and then treated with *n*-octadecene-1 and allowed to stand in hydrogen at 190 °C for 24 h. The Monterey rock shows lower levels of catalytic activity over this reaction period. c, Thermal cracking. Gas product of *n*-octadecene-1 heated at 500 °C in glass for 1 h.

METHODS. Chromatography was carried using a Hewlett-Packard 5980, Series II unit with a 60 m $\times$ 0.53 mm i.d. fused silica capillary column (SPB-1, Supelco) programmed to give 30 °C for 7 min, then to 200 °C at 10 °C min⁻¹. a, Rock samples were first preheated at 350 °C in flowing hydrogen for 24 h to exhaust them of the capacity to generate gas at 190 °C (composition after pretreatment at 350 °C; 10 wt% total organic carbon; Rock-Eval analysis, $S_1 = 0.32$, $S_2 = 34.37$; HI = 335). Rock samples preheated at a lower temperature (250 °C, 24 h) proved inadequate for subsequent catalytic experiments because they retained the capacity to generate gas at 190 °C. Rock samples, of various sizes from a few millimetres to just under a centimetre, were placed in a temperature-controlled glass U-tube (17 ml) fitted at each end with stopcocks and septa and attached to a gas manifold so that hydrogen or helium could flow through. Excess *n*-octadecene-1 (distilled from sodium) was added through the septum to saturate the rock under continued hydrogen flow at 190 °C. Reactions were carried out in closed systems under pure hydrogen at one atmosphere pressure and constant temperature (± 0.5 °C).

1,000 bar are not uncommon in natural gas deposits²³, our experimental conditions are reasonably within the general limits of natural conditions.

Current thinking on the origin of natural gas focuses on the thermal instability of organic matter in sedimentary environments. The distribution of oil and gas in petroleum basins is therefore believed to be linked to thermal history and kerogen type. The gas-to-oil ratio, for example, is thought to increase with depth because oil is believed unstable at the deeper horizons and decomposes to gas^{3,6}. Disturbing contradictions to these views exist, however. For example, oil deposits are found at great depth, at temperatures where only gas should exist³¹ and the light hydrocarbons in natural gas show no evidence of a thermolytic origin³². Our results verify the existence of an alternative, catalytic pathway to natural gas. If it proves to be significant, it could alter the way in which we view the generation and distribution of oil and gas in the Earth. □

suggest even higher temperatures³¹. Pure hydrogen at ambient pressures (~ 1 bar) in our experiments is the same in hydrogen partial pressure as a gas mixture at 1,400 bar containing 700 p.p.m. of hydrogen. As the average composition of natural gas is ~ 700 p.p.m. hydrogen²⁶ and pressures of the order of

Received 13 September 1993; accepted 4 March 1994.

- Schoell, M. *Geochim. cosmochim. Acta* **44**, 649–661 (1980).
- James, A. T. *Bull. Am. Ass. Petrol. Geol.* **67**, 1176–1191 (1983).
- Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence* (Springer, New York, 1984).
- Takach, N. E., Barker, C. & Kemp, M. K. *Bull. Am. Ass. Petrol. Geol.* **71**, 322–333 (1987).
- Barker, C. *Bull. Am. Ass. Petrol. Geol.* **74**, 1254–1261 (1990).
- Ungerer, P. *Org. Geochem.* **16**, 1–25 (1990).
- Evans, R. J. & Felbeck, G. T. Jr *Org. Geochem.* **4**, 135–144 (1983).
- Saxby, J. D. & Riley, K. W. *Nature* **308**, 177–175 (1984).
- Espitalie, J., Ungerer, P., Irwin, I. & Marquis, F. *Org. Geochem.* **13**, 893–899 (1987).
- Horsfield, B., Schenck, H. J., Mills, N. & Welte, D. H. *Org. Geochem.* **19**, 191–204 (1991).
- McNab, J. G., Smith, P. V. & Betts, R. L. *Petrol. Engr. Chem.* **44**, 2556–2563 (1952).
- Fabuss, B. M., Smith, J. O., Lait, R. I., Borsanyi, A. S. & Satterfield, C. N. *Ind. Engr. Chem., Proc. Dev. Div.* **1**, 293–299 (1962).
- Appleby, W. G., Avery, W. H. & Meerbott, W. K. J. *Am. chem. Soc.* **69**, 2279–2285 (1947).
- Mango, F. D. *Geochim. cosmochim. Acta* **56**, 553–555 (1992).
- Laidler, K. J., Sagert, N. H. & Wojciechowski, B. W. *Proc. R. Soc. A* **270**, 242–253 (1962).
- Lewan, M. D., Winters, J. C. & McDonald, J. H. *Science* **203**, 897–899 (1979).
- Tegelaar, E. W. et al. *J. analyt. appl. Pyrolysis* **15**, 29–54 (1989).
- Derenne, S., Largeau, C., Casadevall, E., Berkalo, C. & Rousseau, B. *Geochim. cosmochim. Acta* **55**, 1041–1050 (1991).
- Harwood, R. J. *Bull. Am. Ass. Petrol. Geol.* **61**, 2082–2102 (1977).

- Shock, E. L. *Geology* **16**, 889–890 (1988).
- James, A. T. *Bull. Am. Ass. Petrol. Geol.* **74**, 1441–1458 (1990).
- Fuex, A. N. J. *Geochem. Explor.* **7**, 155–188 (1977).
- Price, L. C. & Wenger, L. M. *Org. Geochem.* **19**, 141–159 (1992).
- Lewan, M. D. *Phil. Trans. R. Soc. A* **331**, 123–134 (1985).
- Monthieux, M., Landais, P. & Monin, J. *Org. Geochem.* **8**, 275–292 (1985).
- Hamak, J. E. & Sigler, S. Information Circular No. 9301 (Bureau of Mines, US Dep. Interior, Pittsburgh, 1991).
- Goebel, E. D., Coveney, R. M. Jr, Angino, E. E., Zeller, E. J. & Dreschhoff, G. A. M. *Oil Gas. J.* 5 May, 215–222 (1984).
- Quigley, T. M. & Mackenzie, A. S. *Nature* **333**, 549–552 (1988).
- Sweeney, J. J., Burnham, A. K. & Braun, R. L. *Bull. Am. Ass. Petrol. Geol.* **71**, 967–985 (1987).
- Price, L. C., Clayton, J. L. & Rumen, L. L. *Trans. Gulf. Coast Ass. Geol. Soc.* **29**, 352–370 (1979).
- Price, L. *Geochim. cosmochim. Acta* **57**, 3261–3280 (1993).
- Mango, F. D. *Nature* **352**, 146–148 (1991).
- Isaacs, C. M. et al. Open-File Report 92–539–C (US Geological Survey, Menlo Park, California, 1992).

ACKNOWLEDGEMENTS. We thank A. Young, L. Price and A. Burnham for their reviews. This work was supported by the US Department of Energy.

Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny

Axel Meyer^{*†}, Jean M. Morrissey^{*}
& Manfred Scharl[‡]

^{*} Department of Ecology and Evolution, and
[†] Program in Genetics, State University of New York,
Stony Brook, New York 11794-5245, USA

[‡] Theodor-Boveri-Institut für Biowissenschaften,
Biozentrum der Universität Würzburg, Am Hubland,
97074 Würzburg, Germany

DARWIN¹ believed that sexual selection accounts for the evolution of exaggerated male ornaments, such as the sword-like caudal fin extensions of male fishes of the genus *Xiphophorus*, that appear detrimental to survival. Swordtails continue to feature prominently in empirical work and theories of sexual selection; the pre-existing bias hypothesis has been offered as an explanation for the evolution of swords in these fishes^{2,3}. Based upon a largely morphological phylogeny, this hypothesis suggests that female preference to mate with sworded males arose in ancestrally swordless species, thus pre-dating the origin of the sword itself and directly driving its evolution. Here we present a molecular phylogeny (based on mitochondrial and nuclear DNA sequences) of *Xiphophorus* which differs from the traditional one: it indicates that the sword originated and was lost repeatedly. Our phylogeny suggests that the ancestor of the genus is more likely to have possessed a sword than not, thus questioning the applicability of the pre-existing bias hypothesis as an explanation for the evolution of this sexually selected trait.

The evolution of exaggerated male secondary sexual characteristics remains one of the most persistent problems in evolutionary biology, and many models of sexual selection have been offered to explain this phenomenon (reviewed in refs 4–7). Some of the most recent models are derived from the 'sensory drive' hypothesis (reviewed in ref. 8), which attempts to explain the origin and evolution of signal design and reception. In the context of sexual selection, the 'sensory exploitation'^{7,9–11} and 'pre-existing bias'^{2,3} hypotheses propose that the evolution of extreme male traits is driven by pre-existing female sensory biases that arose through physiological properties of female sensory systems. Thus, females evolve a sensory bias, which evolutionarily pre-dates and drives the subsequent origin and direction of the evolution of lavish male traits. Therefore, unlike other models of sexual selection, these hypotheses can be tested in a phylogenetic framework.

One of the two commonly cited (see for example refs 12, 13) examples for the pre-existing bias hypothesis involves fishes of the genus *Xiphophorus* (for other examples, see ref. 14). *Xiphophorus* belong to the family Poeciliidae, freshwater fishes from Central America with internal fertilization accomplished by means of the gonopodium, the modified anal fin of males. Males of many species of this genus possess a ventrally elongated caudal fin, the sword. Various definitions of the sword have been used, yet experts on the fish generally regard any extension of the ventral caudal fin rays as a sword (R. Borowsky, K. D. Kallman and M. Scharl, personal communication). Fishes of the genus *Xiphophorus* are commonly divided into swordtails and platyfishes, the latter being formerly placed into a different genus. Although most platyfishes do not possess swords, *X. xiphidium* and *X. andersi* exhibit small swords similar in length or even longer than those of some of the swordtail species (*X. pygmaeus*, *X. continens* and *X. birchmanni*) (Fig. 1).

The present phylogenetic estimate for *Xiphophorus* (Fig. 1) is primarily based on morphological characters, many of which

were derived from the gonopodium and its associated structures¹⁵, but also uses allozyme variation for the northern swordtails¹⁶. In this phylogeny, the platyfishes are paraphyletic; the northern platyfishes are basal to the southern platies, which in turn are more closely related to the swordtails than to the other platies (Fig. 1). Recent phylogenetic work¹⁷, which adds behavioural characters, differs from the previous phylogeny in suggesting that a southern platy, *X. maculatus*, is the most basal member of the genus and that two sworded species, a platy (*X. xiphidium*) and *X. andersi*, group with the swordtails rather than the platies.

Three criteria must be met to show that pre-existing biases have resulted in the evolution of swords *Xiphophorus*². (1) Demonstration that female choice is based on variation in the trait. In choice experiments, female *X. helleri* preferred to associate with males with longer swords over males with shorter swords¹⁸. (2) Evidence must be provided that in species in which the males do not possess the trait, the females prefer to mate with males that do. In choice experiments, female platyfish (*X. maculatus*) spent more time with conspecific males that had long, artificial swords attached (similar to those of *X. helleri*) over naturally swordless conspecific males^{2,3,18}. (3) The absence of the trait must be shown to be the primitive condition. This condition requires phylogenetic knowledge, and appears to be met, given the traditional phylogeny already described (Fig. 1). Because the platies are placed at the base of the genus and are generally swordless, it seems most parsimonious to assume that the common ancestor of the platyfishes and the more derived swordtails was likely to have been a swordless species. Hence, the pre-existing bias hypothesis^{2,3} seems to explain the evolution of swords in *Xiphophorus*.

The applicability of the pre-existing bias hypothesis to *Xiphophorus* critically depends on the accuracy of the phylogeny. If the common ancestor of the genus was likely to have been sworded¹⁶, this hypothesis would not hold because the female preference would not have pre-dated the origin of the sword in

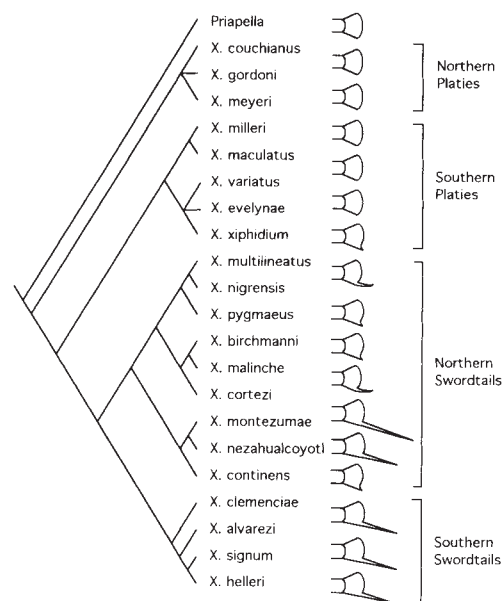


FIG. 1 Traditional morphology-based phylogenetic hypothesis for species of the genus *Xiphophorus*^{15,16} based on ref. 3. The genus *Xiphophorus* has traditionally been divided into four groups (Fig. 1). The phylogenetic position of *X. andersi* has not been determined on the basis of morphological characters, and is therefore not included in this phylogeny. *X. andersi* has been placed as the sister group of the northern platyfishes. *Priapella* is the sister group (its males do not have elongated tails) to *Xiphophorus*^{15,19,34}. This phylogeny suggests that the swordless condition might be ancestral and the sworded condition is derived.

Xiphophorus. In the traditional phylogeny, four of the seven synapomorphies that united swordtails were traits related to the sword¹⁵. As the origin of the sword is of interest, one might want to avoid the danger of circularity and not include characters that are related to the sword in a phylogenetic analysis.

To provide an estimate of the phylogeny of *Xiphophorus* that is independent of the character to be studied, we constructed a DNA-based phylogeny. We included all 22 currently recognized species of the genus, and six outgroup species from the remaining three genera in the tribe Poeciliini¹⁹. The data set consists of 1,284 base pairs (bp) from two partial mitochondrial gene sequences (cytochrome *b* and the control region) and one nuclear gene (the tyrosine kinase gene *X-src*). The data were analysed with the maximum parsimony²⁰ (Fig. 2) and neighbour-joining methods²¹ (Fig. 3). Both methods yielded almost identical phylogenetic estimates which differ, however, from the previous phylogeny (Figs 1, 2 and 3). The molecular and traditional phylogenies agree on the monophyly of the northern platies and of the northern swordtails, and on many relationships within these groups. But the molecular data suggest that the genus *Xiphophorus* is split not into the platies and a derived group composed of northern and southern swordtails, but rather into

two different groups: one consisting of the northern swordtails, and the other of the southern swordtails plus the platies. Within the latter group, the southern swordtails (except *X. clemenciae*, see later) appear to be the sister group of platies. Previously, all southern platies were believed to be a monophyletic group (Fig. 1); in our phylogeny they are paraphyletic, with *Xiphophorus xiphidium* as the sister species of northern platies (Figs 2, 3).

*Xiphophorus andersi*²² has a number of morphological features aligning it to the southern swordtails (for example, it has a small sword and an elongated body) and some linking it to the southern platies (such as gonopodial structure and large head). Its intermediate molecular phylogenetic position as sister species to all platies would seem to confirm its midway morphology (Figs 2, 3). *Xiphophorus clemenciae*, which is traditionally grouped with the southern swordtails (but see ref. 23) because of its close morphological resemblance, is nested among the southern platies instead. Its mitochondrial DNA (mtDNA) does not closely match that of any extant species of *Xiphophorus*; further research will be necessary to test the possibility that this species originated by an ancient hybridization event²⁴. Although hybridization in *Xiphophorus* has not been reported in nature, in labora-

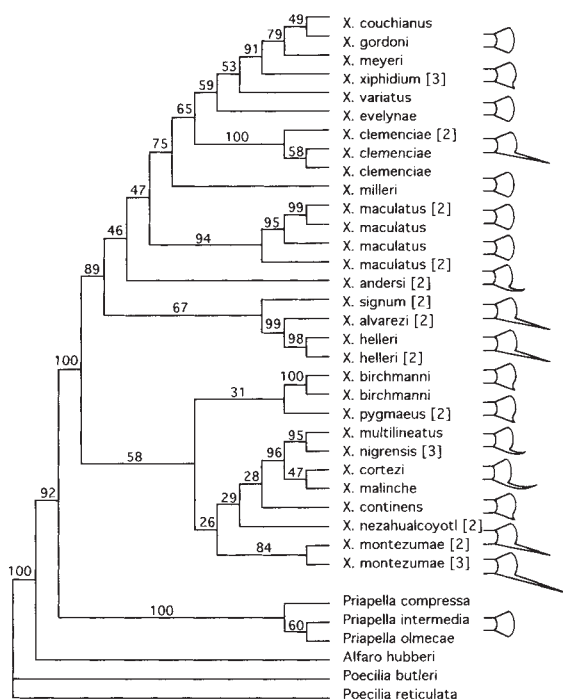


FIG. 2 Fifty per cent majority-rule maximum parsimony bootstrap tree (200 replications) and groupings compatible with it, based on the combined data set of 1,284 bp^{35,20}. We sequenced 52 partial control regions (402 bp), 47 partial cytochrome *b* genes (360 bp), and 42 partial *X-src* genes (522 bp). Sequences have been deposited in Genbank under accession numbers U06488–U06628. DNA sequences were determined after amplification by the polymerase chain reaction (PCR); details of the protocol and the primer sequences used for PCR amplification and direct sequencing have been published (except for the primer L15995 [L-Pro]: 5'-AACTCTACCCCTAGCTCCCAAG-3' for the control region, which was designed for this study)^{36–39}. To improve computational feasibility of this large data set, individuals with sequences that were identical or differed in only one or two bp were combined, and IUPAC symbols were used in the variable sites (number of individuals pooled are given in parentheses after the species name). Thus, the number of taxa was reduced by 16 to 36. In addition to all 22 species of *Xiphophorus*, the analysis included members from all other genera of the tribe Poeciliini¹⁹: *Priapella* (*intermedia*, *compressa*, *olmeae*), *Alfaro hubberi*, and *Poecilia* (*butleri* and *reticulata*, the guppy),

which were declared as the outgroup. Information on sampling localities and collectors for each specimen can be obtained from A.M. No additions or deletions were detected in cytochrome *b*, hence the alignment was unambiguous. The control region and *X-src* were aligned using CLUSTAL⁴⁰, with adjustments by eye. *X-src* is a slowly evolving gene and no amino-acid substitutions were inferred among species of *Xiphophorus*; almost all variation detected was in the intron sequences. The alignments of 28 bp of the control region and 26 bp of *X-src* were questionable; these regions were excluded from the analyses. Data were analysed with maximum parsimony (MP)²⁰ and neighbour-joining (NJ) methods²¹ (Fig. 3), and the robustness of the phylogenetic hypotheses were tested by bootstrapping³⁵. In all parsimony analyses, the following options were used: heuristic search, MULPARS option in effect, GAPMODE = MISSING, and TBR branch swapping. For the control region (no weights), only six shortest trees (length 335) (CI = 0.633, RC = 0.482) were found. For cytochrome *b* (no weights), 81 shortest trees (length 285 steps) (CI = 0.596, RC = 0.438) were found. For *X-src* (no weights), 644 equally short trees were found (length 99) (CI = 0.778, RC = 0.664) were found. The phylogenetic information in the control region sequences seems most reliable because it resulted in the fewest shortest trees. In a regression analysis of all species of *Xiphophorus*, a 2:1 transition-to-transversion ratio (TS:TV) for the control region was observed. In cytochrome *b*, an 8:1 TS:TV was observed within the genus. In *X-src*, transitions and transversions were observed in approximately equal frequencies. In phylogenetic analyses of only *X-src* or cytochrome *b* sequences, the bootstrap values tended to be low; they were much higher in analyses of the control region alone. Cytochrome *b* provided little resolution at the 50% majority rule bootstrap level for deeper nodes, but strongly supported the monophyly of all species and groups of very closely related species. In some separate analyses of cytochrome *b*, the tree is rooted at the northern platies or *X. andersi*. Its topology tended to be susceptible to differential weighting of transitions and transversions. Also, *X-src* provided little resolution at the 50% bootstrap level, which was expected, given the few number of variable sites in this conservative gene. The tree shown is based on equal weights for transitions and transversions in the cytochrome *b* and *X-src* genes, and transversions weighted twice transitions in the control region. The combined data set was robust to changes in weights for transversions in the different genes. Different weights caused only differences in the position of *X. andersi* (branching before or after the *signum*-*alvarezii*-*helleri* clade), the position of *X. xiphidium* in relation to the northern platies, the relation of *cortezi* and *malinche* to the *multilineatus*-*nigrensis* clade, and the respective positions of *continens* and *nezahualcoyotl*. None of the changes due to different weighting regimes in the combined data set affected the reconstruction of the sword in the ancestor of the genus *Xiphophorus* (Fig. 4). If the topology of the traditional phylogenetic hypothesis is forced upon the total molecular data set, the number of evolutionary steps, or tree length, increases by over 10% (569 + versus 510 unweighted MacClade steps with soft polytomies)⁴¹.

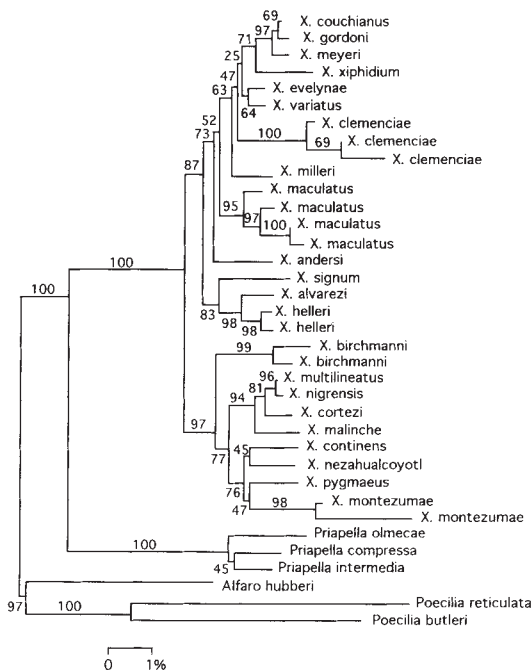


FIG. 3 Majority-rule bootstrap (500 replications) neighbour-joining trees²¹ performed with the program MEGA⁴² on the distance matrix for the complete data set adjusted for multiple hits using Kimura's two-parameter model and pairwise deletion of gaps. Branches are drawn to scale, with the bar representing per cent divergence. The NJ analysis results in a topology that is almost identical to the MP phylogeny (Fig. 2), minor differences being the groupings of *variatus* and *evelynae* as sister species, and relationships among some members of the northern swordtail clade.

tory tests *X. pygmaeus* females prefer to mate with *X. nigrensis* males over conspecific males owing to their more elaborate courtship displays^{25,26}.

In regards to the pre-existing bias theory for *Xiphophorus*, our phylogeny suggests that the ancestor for the entire genus possessed a sword (Fig. 4). Other evidence exists supporting the hypothesis of a sworded ancestor. Through hormone treatments, elongation of the caudal fin can be induced in several species of platyfish tested (for example, in *X. milleri* and *X. maculatus*) which normally do not express this trait^{27–29}. These experiments demonstrate that 'sword' genes and the capability to express them during development are present in some platyfishes without swords. If the molecular phylogeny is correct, this capability might have been retained in these species from ancestrally sworded species (Figs 2, 3 and 4). Sword-like structures occur outside the genus *Xiphophorus*, for example in *Poecilia petenensis*. They are also present in low frequency in natural populations of the guppy, *P. reticulata*, and can be selected for artificially (J. A. Endler, personal communication). Swords might be phylogenetically old, and the capability to express swords might have been present even in the common ancestor of all *Xiphophorus*^{16,30} (Fig. 4).

It appears that swords were lost phylogenetically in some species of platies, despite a continued existence of a female preference for sworded males (in *X. maculatus*, for example²) and the capability of male platies to express caudal fin elongations after hormone treatment. The presence of a sword in males presumably confers a cost in terms of natural selection which might counteract any benefits due to sexual selection by increased mating success. Males with swords may suffer a decrease in swimming speed and an increase in visibility, resulting in an increased risk of predation. The potential costs in survivorship may have led to the loss of swords, despite a mating advantage of males with swords over males without them.

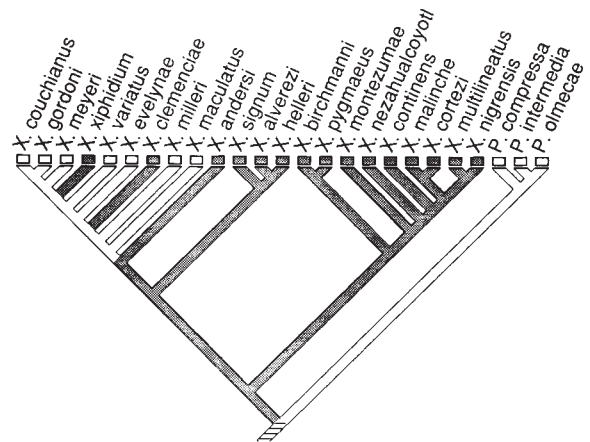


FIG. 4 Mapping of the traits 'no sword' and 'sword', and most parsimonious reconstruction of the ancestral states on the molecular phylogeny (Fig. 2) using MacClade⁴¹. The reconstruction of the evolution of the sword is based on the MP majority-rule bootstrap phylogeny; the minor differences between the MP and NJ analyses (Figs 2, 3) have no effect on the reconstruction of the sword in the ancestor of the genus. As all species were found to be monophyletic, they are represented by only one terminal taxon. If the presence or absence of the sword is treated as a two-state character (sword–no sword), then the common ancestor most likely possessed a sword (4 steps). Forcing the ancestor of the genus to be swordless adds one more step to the evolution of the sword. Basolo³ defines several characters from the general character complex 'sword' and the sword itself² as 'a coloured extension of the lower margin of the caudal fin'. Under her definition a sword is a coloured ventral extension 0.7–6.0 times the length of the caudal fin, and a protrusion is a coloured ventral extension, 0.1 to 0.3 times the length of the caudal fin. If the characters are coded in this manner with regard to relative length (*X. birchmanni*, *X. pygmaeus*, *X. continens* and *X. xiphidium* have a protrusion) and treated as 'ordered' (that is, protrusion is an intermediate step in the evolution of swords from no swords, NS–P–S, as suggested in ref. 41), then the common ancestor of all species of *Xiphophorus* most probably had a protrusion or a sword rather than no sword (9 steps). Under these conditions of reconstruction, 10 steps would be required if the ancestor was swordless. If the character evolution is assumed to be unordered (with three states), the common ancestor is likely to have had a sword (6 steps); forcing the ancestor to be swordless or to possess a protrusion adds a step to the character reconstruction (7 steps).

Female preferences and preferred male traits might be expected to be genetically linked³¹ and their positive genetic correlation has been demonstrated^{32,33}. The evolutionary dissociation of swords and female preference therefore seems odd. Based on the molecular phylogeny, it appears as if caudal fin extensions can be lost and regained (for example, in *X. xiphidium*) (Fig. 4). Basolo's² finding of a preference for swords in a species whose males are lacking swords is intriguing; but in light of the new phylogenetic hypothesis, the puzzle may not be why males in most species of platyfish never evolved swords despite a female preference for them, but rather why swords were lost in the presence of an apparently retained bias in females to mate with males with swords. □

Received 17 November 1993; accepted 17 January 1994.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
2. Basolo, A. L. *Science* **250**, 808–810 (1990).
3. Basolo, A. L. *Science* **253**, 1426–1427 (1991).
4. Bradbury, J. W. & Anderson, M. B. (eds) *Sexual Selection: Testing the Alternatives* (Wiley, Chichester, UK, 1987).
5. Maynard Smith, J. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Anderson, M. B.) 9–20 (Wiley, Chichester, UK, 1987).
6. Maynard Smith, J. *Trends Ecol. Evol.* **6**, 146–151 (1991).
7. Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33–38 (1991).
8. Endler, J. A. *Am. Nat.* **139**, S125–S153 (1992).
9. Kirkpatrick, M. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Anderson, M. B.) 67–82 (Wiley, Chichester, UK, 1987).
10. Ryan, M. *Oxf. Surv. Evol. Biol.* **7**, 157–195 (1990).

11. Ryan, M. J., Fox, J. H., Wilczynski, W. & Rand, A. S. *Nature* **343**, 66–68 (1990).
12. Alcock, J. *Animal Behavior* 5th edn (Sinauer, Sunderland, 1993).
13. Krebs, J. R. & Davies, N. B. *An Introduction to Behavioral Ecology* 3rd edn (Blackwell, Oxford, 1993).
14. Ryan, M. & Rand, A. S. *Phil. Trans. R. Soc. B* **340**, 187–195 (1993).
15. Rosen, D. E. *Bull. Am. Mus. nat. Hist.* **182**, 268–375 (1979).
16. Rauchenberger, M., Kallman, K. D. & Morizot, D. C. *Am. Mus. nat. Hist. Nov.* **2975**, 1–41 (1990).
17. Haas, V. in *Trends in Ichthyology* (eds Schröder, J. H., Bauer, J. & Scharlt, M.) 279–288 (Blackwell, London, 1993).
18. Basolo, A. L. *Anim. Behav.* **40**, 332–338 (1990).
19. Rosen, D. E. & Bailey, R. M. *Bull. Am. Mus. nat. Hist.* **126**, 1–176 (1963).
20. Swofford, D. L. *Phylogenetic Analysis Using Parsimony* Version 3.1.1 (Illinois Natural History Survey, Champaign, 1993).
21. Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406–425 (1987).
22. Meyer, M. K. & Scharlt, M. *Senckenbergiana biol.* **60**, 147–151 (1980).
23. Morizot, D. C. & Siciliano, M. J. *Genetics* **102**, 539–556 (1982).
24. Dowling, T. E. & DeMarais, B. D. *Nature* **362**, 444–446 (1993).
25. Ryan, M. J. & Wagner, W. E. *Jr Science* **236**, 595–597 (1987).
26. Partridge, L. & Harvey, P. H. *Nature* **328**, 377 (1987).
27. Gordon, M., Cohen, H. & Nigrelli, R. F. *Am. Nat.* **77**, 569–572 (1943).
28. Dzwillo, M. *Verh. Dt. Zool. Ges.* **1962**, 152–159 (1963).
29. Zander, C. D. & Dzwillo, M. *Z. Wissen. Zool.* **178**, 276–315 (1969).
30. Winquist, S. T. et al. *Science* **253**, 1426 (1991).
31. Fisher, R. A. *The General Theory of Natural Selection* 2nd edn (Dover, New York, 1958).
32. Houde, A. & Endler, J. A. *Science* **248**, 1405–1408 (1990).
33. Bakker, T. C. M. *Nature* **363**, 255–257 (1993).
34. Rosen, D. E. & Kallman, K. D. *Am. Mus. nat. Hist. Nov.* **2379**, 1–29 (1969).
35. Felsenstein, J. *Evolution* **39**, 783–791 (1985).
36. Kocher, T. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6196–6200 (1989).
37. Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. *Nature* **347**, 550–553 (1990).
38. Meyer, A. & Lydeard, C. *Proc. R. Soc. Lond. B* **254**, 153–162 (1993).
39. Robertson, S. M. thesis, Ludwigs Maximilians Univ., München (1989).
40. Higgins, D. G. & Sharp, P. M. *CABIOS* **5**, 151–153 (1989).
41. Maddison, W. P. & Maddison, D. R. *MacClade Ver. 3.0. Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, 1992).
42. Kumar, S., Tamura, K. & Nei, M. *MEGA Molecular Evolutionary Genetics Analysis Vers. 1.0* (Inst. of Mol. Evol. Genetics, Pennsylvania State Univ., 1993).
43. Meyer, M. K. *Cour. Forsch.-Inst. Senckenberg* **94**, 123–130 (1987).

ACKNOWLEDGEMENTS. We thank R. Borowsky, J. Endler, R. McKeand and particularly K. Kallmann for specimens; S. Robertson for X-src primer design; C. Davis for technical assistance; and A. Basolo, R. Borowsky, J. Endler, K. Kallmann, C. Lydeard, E. Mayr, E. Prager and M. Ryan for comments on the manuscript. Part of this work was completed while A.M. was an Alfred P. Sloan Postdoctoral Fellow in the laboratory of the late Allan C. Wilson at the University of California at Berkeley. A.M. is supported by the US NSF; M.S. is supported by the DFG, Germany.

The generic viewpoint assumption in a framework for visual perception

William T. Freeman

Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, Massachusetts 02139, USA

A VISUAL system makes assumptions in order to interpret visual data. The assumption of 'generic view'^{1–4} states that the observer is not in a special position relative to the scene. Researchers commonly use a binary decision of generic or accidental view to disqualify scene interpretations that assume accidental viewpoints^{5–10}. Here we show how to use the generic view assumption, and others like it, to quantify the likelihood of a view, adding a new term to the probability of a given image interpretation. The resulting framework better models the visual world and reduces the reliance on other prior assumptions. It may lead to computer vision algorithms of greater power and accuracy, or to better models of human vision. We show applications to the problems of inferring shape, surface reflectance properties, and motion from images.

Consider the image of Fig. 1a. Perceptually, there are two possible interpretations: a bump, lit from the left, or a dimple, lit from the right. Yet many shapes and lighting directions (Fig. 1b) could explain the image. How should a visual system choose?

We note that the ridges in shapes 2–4 of Fig. 1b must line up with the assumed light direction. We can study the 'accidentalness' of this alignment by exploring how the image of the illuminated shape changes as we perturb the azimuthal light direction. Figure 1c shows that shape 3 presents images similar to that in

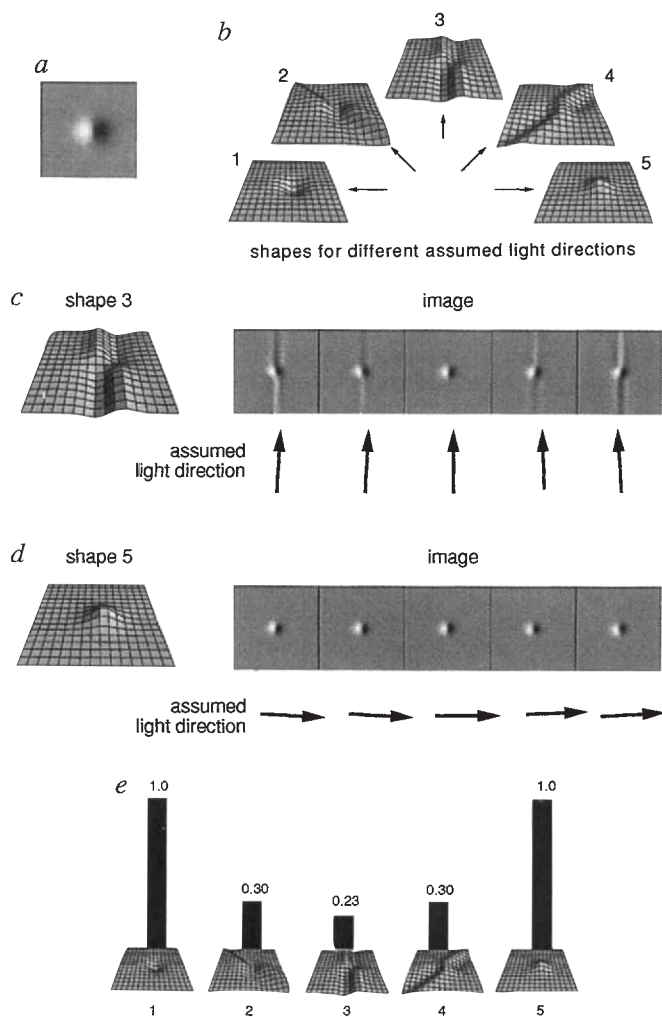


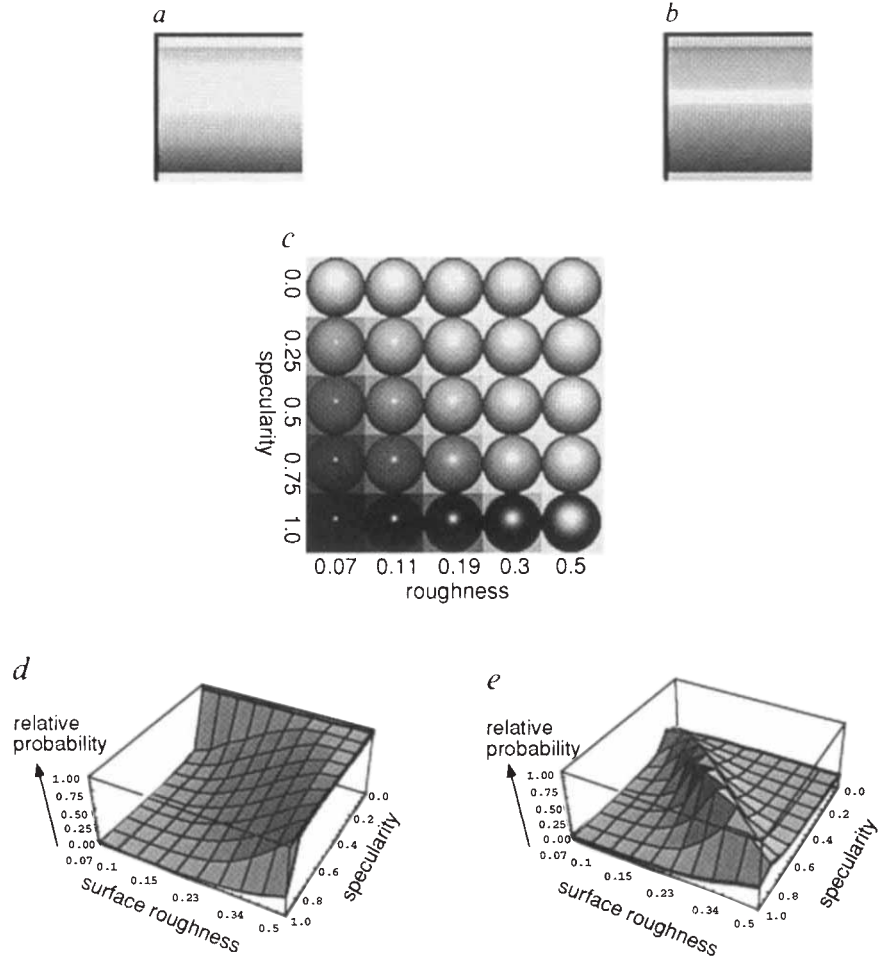
FIG. 1 a, Perceptually, this image has two possible interpretations. It could be a bump, lit from the left, or a dimple, lit from the right. b, Mathematically, there are many possibilities. The five shown here were found by a linear shape from shading algorithm, assuming shallow incident light from different azimuthal directions and the boundary conditions described in ref. 8. Shapes 2–4 require coincidental alignment with the assumed light direction. For shape 3 in c, the rendered image changes quickly with assumed light angle; only a small range of light angles yields an image like that shown in a. The generic view term of the scene probability equation, equation (7), penalizes an interpretation that has high image derivatives with respect to the generic variable, in this case light direction. For shape 5 in d, a much larger range of light angles gives the observed image. If all light directions are equally likely, shape 5 should be the preferred explanation. The probabilities of the candidate shapes, found using equation (7), are shown in e. The results favour shapes 1 and 5, in agreement with the perceptual appearance of a.

Fig. 1a only for a small range of assumed light directions. The bump in Fig. 1d (shape 5) presents images like that in Fig. 1a over a broader range of light directions. If all azimuthal light directions are equally likely, shape 5 has more chances to create the image in Fig. 1a than does shape 3.

To quantify such probabilities, we use a bayesian framework (as in ref. 11, for example). This combines the data (Fig. 1a) with known or estimated prior probabilities to find the posterior probability of each candidate shape.

We treat the azimuthal light direction as a random variable, an example of what we call a generic variable, $\bar{x}$, with prior probability density $P_{\bar{x}}(\bar{x})$. (We use subscripts to distinguish

FIG. 2 *a* and *b*, Two images with intensity variations along only one dimension. Such images can be explained by many different combinations of surface reflectance function and shape. We use a two-parameter family of reflectance functions (a subset of the model of ref. 21) and a fixed light position to generate a family of possible shape and reflectance function explanations for each of *a* and *b*. *c*, Visual key to the parameters provided by showing the appearance of the surface reflectance functions, rendered on the surface of a sphere. For every specularity and roughness, shapes exist that produce image *a* or *b*. (For each shape we assumed boundary conditions of constant height at the vertical picture edge.) One wants to choose between these competing explanations without resorting to an *ad hoc* bias toward some shapes or reflectance functions. Each of the explanations will present the images shown over differing ranges of the generic variables, taken here to be vertical light angle and object orientation. The scene probability equation calculates their relative probability densities²². Plots *d* and *e* show the probability that the images *a* and *b*, respectively, were created by each surface reflectance function in the parameter space and corresponding shape. The probabilities are the highest for the reflectance functions that look like the material of the corresponding original image (compare with *c*).



between probability densities, P .) Generic variables can include viewpoint, lighting direction, or object pose. These are variables that we do not need to estimate precisely.

We assume a prior probability density, $P_{\beta}(\tilde{\beta})$, for the scene parameter $\tilde{\beta}$ we want to estimate. For this example, shapes 1–5 are assigned equal probabilities. The posterior distribution, $P(\tilde{\beta}, \tilde{x} | \tilde{y})$, gives the probability that scene parameter $\tilde{\beta}$ (shape) and generic variable $\tilde{x}$ (light direction) created the visual data $\tilde{y}$ (Fig. 1a). From $P(\tilde{\beta}, \tilde{x} | \tilde{y})$, we will find the posterior probability $P(\tilde{\beta} | \tilde{y})$.

We use Bayes' theorem to evaluate $P(\tilde{\beta}, \tilde{x} | \tilde{y})$:

$$P(\tilde{\beta}, \tilde{x} | \tilde{y}) = \frac{P(\tilde{y} | \tilde{\beta}, \tilde{x}) P_{\beta}(\tilde{\beta}) P_{\tilde{x}}(\tilde{x})}{P_{\tilde{y}}(\tilde{y})} \quad (1)$$

where we have assumed that $\tilde{x}$ and $\tilde{\beta}$ are independent. The denominator is constant for all models $\tilde{\beta}$ to be compared.

To find $P(\tilde{\beta}, \tilde{x} | \tilde{y})$ independently of the value of the generic variable $\tilde{x}$, we integrate the joint probability of equation (1) over the possible $\tilde{x}$ values:

$$P(\tilde{\beta} | \tilde{y}) = \frac{P_{\beta}(\tilde{\beta})}{P_{\tilde{y}}(\tilde{y})} \int P(\tilde{y} | \tilde{\beta}, \tilde{x}) P_{\tilde{x}}(\tilde{x}) d\tilde{x} \quad (2)$$

We will assume that the prior probability $P_{\tilde{x}}(\tilde{x})$ of the generic variables is a constant. The generalization for other priors is straightforward. $P(\tilde{y} | \tilde{\beta}, \tilde{x})$ is large where the scene $\tilde{\beta}$ and the value $\tilde{x}$ give an image similar to the observation $\tilde{y}$. The integral of equation (2) integrates the area of $\tilde{x}$ for which $\tilde{\beta}$ yields the observation. In our example, it effectively counts the frames in Fig. 1c or d, where the rendered image is similar to the input data.

We assume zero mean gaussian observation noise of variance σ^2 , which plays two roles. It measures the similarity between images as the probability that noise accounts for the differences. It can also model physical noise. For this noise model,

$$P(\tilde{y} | \tilde{\beta}, \tilde{x}) = \frac{1}{(\sqrt{2\pi\sigma^2})^N} e^{-\|\tilde{y} - \tilde{f}(\tilde{x}, \tilde{\beta})\|^2 / 2\sigma^2} \quad (3)$$

where $\tilde{f}(\tilde{x}, \tilde{\beta})$ is a known 'rendering function' which gives the image created by the generic and scene parameters $\tilde{x}$ and $\tilde{\beta}$, and N is the dimensionality of the visual data $\tilde{y}$.

For the low noise limit, we can find an analytic approximation to the integral of equation (2). We expand $\tilde{f}(\tilde{x}, \tilde{\beta})$ in equation (3) in a second-order Taylor series,

$$\tilde{f}(\tilde{x}, \tilde{\beta}) \approx \tilde{f}(\tilde{x}_0, \tilde{\beta}) + \sum_i \tilde{f}'_i [\tilde{x} - \tilde{x}_0]_i + \frac{1}{2} \sum_{i,j} [\tilde{x} - \tilde{x}_0]_i \tilde{f}''_{ij} [\tilde{x} - \tilde{x}_0]_j \quad (4)$$

where $[\cdot]_i$ indicates the i th component of the vector in brackets, and

$$\tilde{f}'_i = \left. \frac{\partial \tilde{f}(\tilde{x}, \tilde{\beta})}{\partial x_i} \right|_{\tilde{x} = \tilde{x}_0} \quad (5)$$

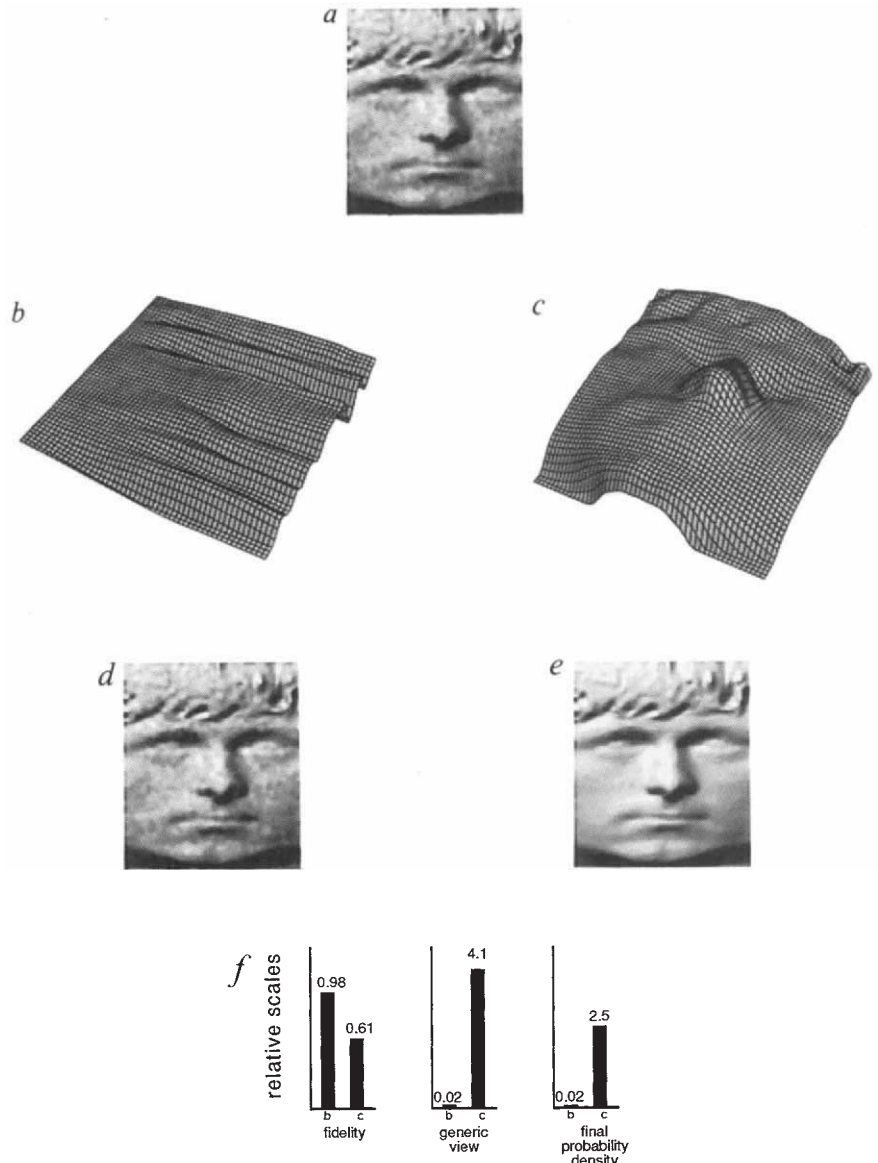
and

$$\tilde{f}''_{ij} = \left. \frac{\partial^2 \tilde{f}(\tilde{x}, \tilde{\beta})}{\partial x_i \partial x_j} \right|_{\tilde{x} = \tilde{x}_0} \quad (6)$$

We take $\tilde{x}_0$ to be the value of $\tilde{x}$ which can best account for the observed image data; that is, for which $\|\tilde{y} - \tilde{f}(\tilde{x}, \tilde{\beta})\|^2$ is minimized.

Using equations (3)–(6) to second order in $\tilde{x} - \tilde{x}_0$ in the integral of equation (2), we find the posterior probability for the

FIG. 3 Showing the need for the generic view term of equation (7). We compare the probability densities of two explanations for the image in *a*. The surface in *b* (shown at $7\times$ vertical exaggeration), lit at a grazing angle, yields the image in *d*. The surface in *c* gives the image in *e*, which accounts less well for the image in *a*. Thus, based on an image fidelity criterion, *b* is a better explanation. The common prior assumption of a smooth surface¹⁴ would also favour *b* (the surface is very smooth at the true vertical scale). However, the object and light source must be precisely positioned for the shape in *b* to give the image in *d*; the generic view term of the scene probability equation (7) penalizes this. Including the generic view term makes the overall probability densities (shown in *f*), favour the perceptually reasonable explanation of shape *c* over shape *b*. (We made this example by construction. Gaussian random noise at a 7 dB signal-to-noise ratio was added to *e* to make *a*; *b* was found from *a* using a shape from shading algorithm, assuming constant surface height at the left picture edge²³. We evaluated the likelihood of *b* and *c*, assuming both generic object pose and generic lighting direction²⁴. The strength of a prior preference for smooth surfaces is arbitrary and none was included in the final densities. The actual noise variance was used for σ^2 in the fidelity term of equation (7), although a wide range of assumed variances would give the preferences shown here.)



scene parameters $\tilde{\beta}$ given the visual data $\tilde{y}$:

$$P(\tilde{\beta} | \tilde{y}) = k \exp \left(\frac{-\|\tilde{y} - \tilde{f}(\tilde{x}_0, \tilde{\beta})\|^2}{2\sigma^2} \right) P_{\beta}(\tilde{\beta}) \frac{1}{\sqrt{\det(\mathbf{C})}} \quad (7)$$

= *k* (fidelity) (prior probability) (generic view)

where the *i* and *j*th elements of the matrix **C** are

$$C_{ij} = \tilde{f}'_i \cdot \tilde{f}'_j - (\tilde{y} - \tilde{f}(\tilde{x}_0, \tilde{\beta})) \cdot \tilde{f}''_{ij} \quad (8)$$

We call equation (7) the scene probability equation. The normalization constant *k* does not enter into comparisons between interpretations $\tilde{\beta}$. The exponential term, which we call the image fidelity term, favours scene hypotheses that have a small mean-squared difference from the visual data. This and the prior probability term $P_{\beta}(\tilde{\beta})$ are familiar in computational vision. Regularization, from which many vision algorithms have been derived^{12,13}, finds the maximum probability density^{14,15} using these two terms, when viewed in a bayesian context. The third, generic view term, accounts for the assumptions of generic viewpoint, pose or lighting position. The scene probability equation favours interpretations that can generate the observed image over a relatively large range of generic variables, by penalizing

large image derivatives with respect to those variables. If the prior probability of the generic variable were not constant, then the factor $P_{\beta}(\tilde{x}_0)$ would be included in the prior term of equation (7).

The generic view term is especially useful when several explanations account equally well for visual data, as occurs commonly in problems of stereo, shape, motion and colour perception; ref. 16, for example. Then the image fidelity term is the same for the competing explanations. The prior probabilities may not be known well⁴. The generic view term allows a choice based on the relatively reliable assumptions of generic view, pose, or light source position.

Our approach relates to bayesian analyses of data interpolation, image restoration and other problems^{11,15,17}. In that work, as in this, one favours hypotheses that could have generated the observed data in many ways (see also ref. 18, a related non-bayesian approach).

Using the scene probability equation (7), we plot in Fig. 1e the relative probabilities of shapes 1–5 of Fig. 1b. Note the agreement with the bump/dimple shapes perceived to be the true explanation of Fig. 1a. (Presumably, these are perceptually favoured because they are more probable.) Without the generic

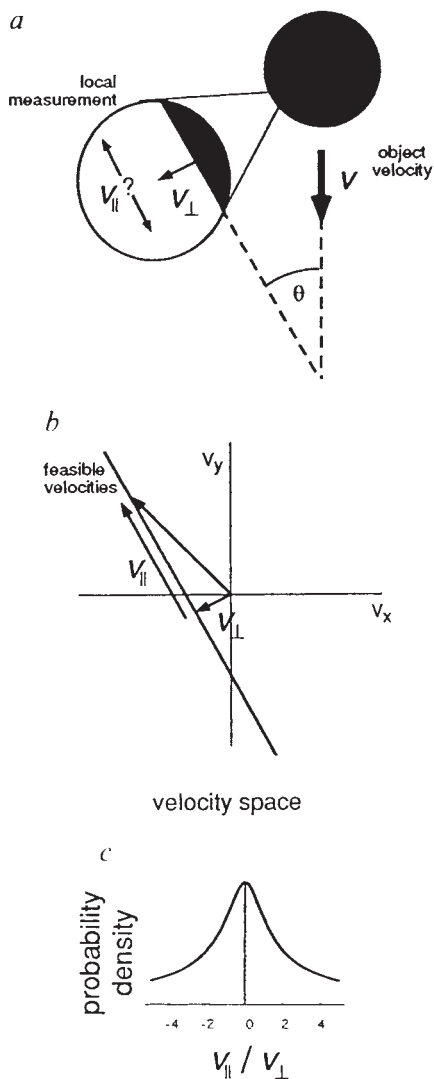


FIG. 4 Application of the scene probability equation to velocity estimation. *a*, Within a local aperture, the object velocity direction is ambiguous¹⁹. $V_{\perp}$, the component of velocity normal to the local contrast, is constrained by the measurement, while $V_{\parallel}$ is unconstrained. *b*, Line in velocity space of object velocities consistent with observed normal velocity. High values of $V_{\parallel}$ imply a coincidental alignment of the local contrast orientation with the object velocity direction. In our framework, the measurement vector $\bar{y}$ is the normal velocity vector; the scene parameter β is $V_{\parallel}$; the generic variable $\bar{x}$ is the angle θ between the object velocity and the orientation of local contrast. The scene probability equation (7), penalizes high derivatives of the normal velocity with respect to contrast orientation. *c*, Resulting posterior probability for $V_{\parallel}$, showing a bias in favour of the normal velocity ($V_{\parallel} = 0$). This bias is consistent with psychophysical observations²⁰.

view term, one would have to state an arbitrary preference for bumps or dimples to choose between the candidate shapes.

In Fig. 2 we use the scene probability equation to choose between surface reflectance functions in a case where they would otherwise be indistinguishable. Figure 3 shows an example in which both the fidelity and the prior probability terms favour a perceptually implausible explanation. Only when the generic view term of equation (7) is included does the perceptually favoured explanation rank higher.

In Fig. 4, we apply the scene probability equation to the problem of estimating the local image velocity from local measurements of the velocity components normal to the contrast orientation¹⁹. All velocity components parallel to the local con-

trast orientation are possible, but high speeds would imply a coincidental alignment of the local contrast with the image velocity. The scene probability equation predicts a bias toward zero parallel velocity component, which is supported by psychophysical evidence²⁰.

From an equation that ranks scene interpretations, such as the scene probability equation (equation (7)), one can develop vision algorithms that find an optimum interpretation. Including the generic view term gives a better statistical model of the visual world. It may result in more powerful and accurate algorithms for vision. □

Received 13 September 1993; accepted 27 January 1994.

1. Koenderink, J. J. & van Doorn, A. J. *Biol. Cybern.* **32**, 211–216 (1979).
2. Binford, T. O. *Art. Intel.* **17**, 205–244 (1981).
3. Biederman, I. *Comp. Vis. Graph. Image Proc.* **32**, 29–73 (1985).
4. Nakayama, K. & Shimojo, S. *Science* **257**, 1357–1363 (1992).
5. Lowe, D. G. & Binford, T. O. *IEEE Pat. Anal. Mach. Intel.* **7**, 320–326 (1985).
6. Malik, J. *Intl. J. Comp. Vis.* **1**, 73–103 (1987).
7. Richards, W. A., Koenderink, J. J. & Hoffman, D. D. *J. Opt. Soc. Am. A*, **4**, 1168–1175 (1987).
8. Pentland, A. P. *Intl. J. Comp. Vis.* **1**, 153–162 (1990).
9. Leclerc, Y. G. & Bobick, A. F. in *Proc. IEEE Computer Vision and Pattern Recognition* 552–558 (Maui, Hawaii, 1991).
10. Jepson, A. D. & Richards, W. *Spatial Vision in Humans and Robots* (eds Harris, L. & Jenkin, M.) (Cambridge Univ. Press, UK, 1992).
11. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis* (Springer, New York, 1985).
12. Tikhonov, A. N. & Arsenin, V. Y. *Solutions of Ill-posed Problems* (Winston, Washington DC, 1977).
13. Poggio, T., Torre, V. & Koch, C. *Nature* **317**, 314–319 (1985).
14. Terzopoulos, D. *IEEE Pat. Anal. Mach. Intel.* **8**, 413–424 (1986).
15. Szeliski, R. *Bayesian Modeling of Uncertainty in Low-level Vision* (Kluwer Academic, Boston, 1989).
16. Marr, D. C. *Vision* (Freeman, New York, 1982).
17. MacKay, D. J. C., *Neural Comput.* **4**, 415–447 (1992).
18. Weinshall, D., Werman, M. & Tishby, N. *Proc. 3rd Eur. Conf. Computer Vision* (Springer, Stockholm, 1994).
19. Horn, B. K. P. & Schunk, B. G. *Artif. Intel.* **17**, 185–203 (1981).
20. Nakayama, K. & Silverman, G. H. *Vision Res.* **739**–746 (1988).
21. Cook, R. L. & Torrance, K. E. in *SIGGRAPH-81* (Special Interest Group on Computer Graphics Conf. Proc. Assoc. for Computing Machinery, New York, 1981).
22. Freeman, W. T. in *Proc. 4th Intl. Conf. Computer Vision* 347–356 (IEEE, Berlin, Germany, 1993).
23. Bichsel, M. & Pentland, A. P. in *Proc. IEEE Computer Vision and Pattern Recognition* 459–465 (Champaign, IL, 1992).
24. Freeman, W. T. *Intl. J. Comp. Vis.* (in the press).

ACKNOWLEDGEMENTS. I thank E. Adelson, D. Brainard, D. Knill, K. Nakayama, E. Simoncelli, R. Szeliski and A. Yuille for discussion and suggestions. Much of this research was performed at the MIT Media Laboratory and was supported by a contract from David Sarnoff Research Laboratories (subcontract to the National Information Display Laboratory) to E. Adelson.

Direct modulation by Ca^{2+} – calmodulin of cyclic nucleotide-activated channel of rat olfactory receptor neurons

Tsung-Yu Chen & King-Wai Yau

Department of Neuroscience and Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

OLFACTORY receptor neurons depolarize in response to odorant stimulation of their sensory cilia^{1–3}. One transduction mechanism involves a G-protein-mediated increase in adenylate cyclase activity^{4–8}, raising the internal cyclic AMP concentration to open a cyclic nucleotide-activated cation channel on the plasma membrane^{9–14}. An influx of Ca^{2+} through this channel, which is permeable to both monovalent and divalent cations, triggers olfactory adaptation¹⁵. Previous work has indicated that at least part of this Ca^{2+} -mediated adaptation resides in the channel itself^{15–17}, but the mechanism remains unclear and controversial^{16–18}. Here we use the cloned channel from rat¹⁹ expressed in a cell line and the native channel from rat olfactory receptor cells to show that Ca^{2+} reduces the apparent affinity of the channel for cAMP by

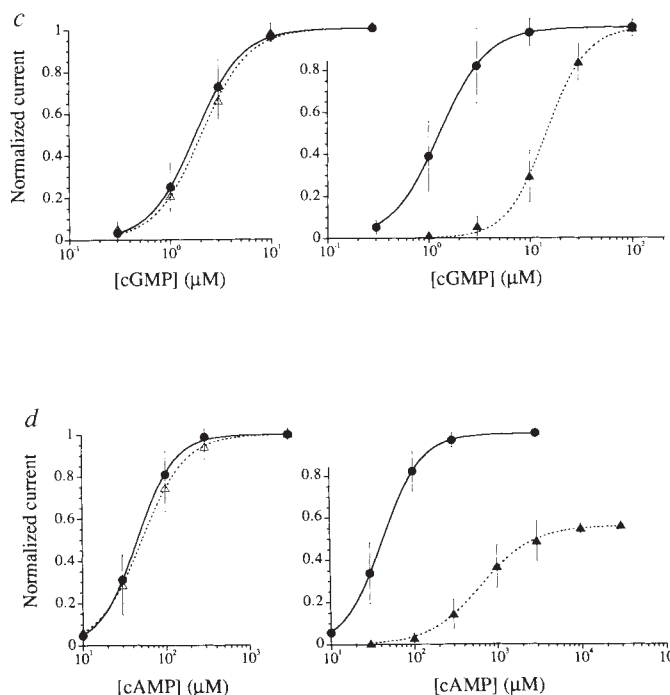
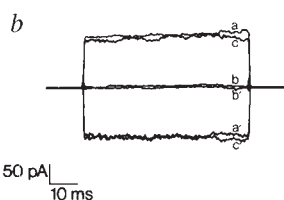
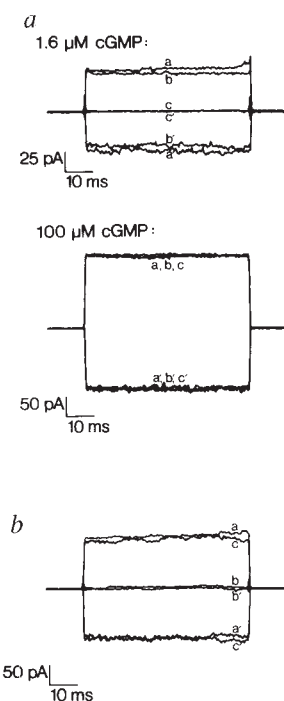
up to 20-fold in the presence of calmodulin, an abundant protein in olfactory cilia²⁰. This decrease in apparent affinity appears to involve a direct interaction between Ca^{2+} -calmodulin and the channel, and it can reduce the activation of the channel by cAMP by up to a few hundred-fold, suggesting that it may be a key component of the Ca^{2+} -triggered olfactory adaptation.

The cAMP/cGMP-activated cation channel complementary DNA cloned from rat olfactory epithelium¹⁹ was transfected into HEK293 cells (see Fig. 1 legend), and patch-clamp recordings were made from excised inside-out patches of plasma membrane of transfected cells^{19,21}. In Fig. 1a (top), the membrane potential of an excised patch with expressed channels was stepped from zero to ± 60 mV under three different bath-solution conditions: (1) zero Ca^{2+} (traces a and a'); (2) $50 \mu\text{M}$ Ca^{2+} (traces b and b'); and (3) $50 \mu\text{M}$ Ca^{2+} and 235 nM calmodulin (traces c and c'), all with $1.6 \mu\text{M}$ cGMP. Ca^{2+} by itself had only a slight effect on the cGMP-induced current, but in the presence of calmodulin it reduced the current to almost zero. This current inhibition

seemed to result from a reduced affinity of the channel for cGMP, because at $100 \mu\text{M}$ cGMP Ca^{2+} -calmodulin had no effect (Fig. 1a lower panel; traces c and c'). The current reduction at $1.6 \mu\text{M}$ cGMP was blocked by the calmodulin inhibitor mastoparan²² (Fig. 1b, traces c and c'), consistent with a calmodulin action. We also tried calmidazolium¹⁶, another calmodulin inhibitor, but this inhibited the channel independent of calmodulin.

Averaged results from several experiments at -60 mV were plotted as dose-response relationships (Fig. 1c). The results at 0 and $50 \mu\text{M}$ Ca^{2+} are almost identical, but, in the presence of 235 nM calmodulin, Ca^{2+} shifts the $K_{1/2}$ value of the dose-response relation by 11-fold. The $K_{1/2}$ shift at $+60$ mV was identical (data not shown), indicating that this effect was not voltage-dependent. With cAMP as ligand (Fig. 1d), the $K_{1/2}$ at -60 mV was increased 15-fold by Ca^{2+} and calmodulin. In addition, Ca^{2+} -calmodulin reduced the maximum cAMP-induced current by about half (see figure legend), but this was not observed in

FIG. 1 a, Effect of Ca^{2+} -calmodulin on the cloned rat olfactory cAMP/cGMP-activated channel at $1.6 \mu\text{M}$ and $100 \mu\text{M}$ cGMP. Traces show currents evoked by a voltage pulse from 0 to ± 60 mV in the presence of bath cGMP and under each of three bath-solution conditions: 0 Ca^{2+} (traces a, a'), $50 \mu\text{M}$ Ca^{2+} (b, b'), and $50 \mu\text{M}$ Ca^{2+} with 235 nM calmodulin (c, c'). Leakage current in the absence of cGMP has been subtracted. b, Blockage of the Ca^{2+} -calmodulin effect by the calmodulin inhibitor mastoparan. The solution contained $1.6 \mu\text{M}$ cGMP, together with $50 \mu\text{M}$ Ca^{2+} (traces a, a'), $50 \mu\text{M}$ Ca^{2+} with 235 nM calmodulin (b, b'), or $50 \mu\text{M}$ Ca^{2+} with 235 nM calmodulin and $1.2 \mu\text{M}$ mastoparan (c, c'). c, Collected dose-response relations between activation of the cloned channel and cGMP concentration in 0 Ca^{2+} (circles), $50 \mu\text{M}$ Ca^{2+} (open triangles), and $50 \mu\text{M}$ Ca^{2+} with 235 nM calmodulin (filled triangles), respectively. For a given patch, only the control (0 Ca^{2+}) and one test condition ($50 \mu\text{M}$ Ca^{2+} with or without calmodulin) could be compared, which explains why there are two control dose-response relations. Current was measured at -60 mV and normalized with respect to the value at $100 \mu\text{M}$ cGMP in zero Ca^{2+} . Points represent averaged measurements from several patches; error bars indicate standard deviations. Fitted curves are drawn from the Hill equation, $I_{\text{norm}} = C^n / (C^n + K_{1/2}^n)$, where I_{norm} is the normalized current, C is the cyclic nucleotide concentration and n is the Hill coefficient. Left panel: 3 patches, with saturating current at -60 mV ranging 62–176 pA; curve fitted with $K_{1/2} = 1.8 \mu\text{M}$, $n = 1.9$ for zero Ca^{2+} (solid) and $K_{1/2} = 2.0 \mu\text{M}$, $n = 1.9$ for $50 \mu\text{M}$ Ca^{2+} (dashed). Right panel: 7 patches, with saturating current at -60 mV ranging 68–676 pA; curve fitted with $K_{1/2} = 1.3 \mu\text{M}$, $n = 1.8$ for zero Ca^{2+} and $K_{1/2} = 14.6 \mu\text{M}$, $n = 2.2$ for $50 \mu\text{M}$ Ca^{2+} with 235 nM calmodulin. d, Same experiment as in c, but with cAMP as ligand. Current was again measured at -60 mV and normalized with respect to value at 3 mM cAMP in zero Ca^{2+} ; symbols as in c. Left panel: 4 patches, with saturating current at -60 mV ranging 85–376 pA; curve fitted with $K_{1/2} = 45.9 \mu\text{M}$, $n = 1.9$ for zero Ca^{2+} (solid) and $K_{1/2} = 52.0 \mu\text{M}$, $n = 1.6$ for $50 \mu\text{M}$ Ca^{2+} (dashed). Right panel: 6 patches, with saturating current ranging 119–3,483 pA; curve fitted with $K_{1/2} = 43.4 \mu\text{M}$, $n = 1.8$ for zero Ca^{2+} and $K_{1/2} = 645 \mu\text{M}$, $n = 1.4$ (with an amplitude scaling factor of 0.56) for $50 \mu\text{M}$ Ca^{2+} with 235 nM calmodulin. The much larger control $K_{1/2}$ value for cAMP compared with



that for cGMP has been found before for the expressed channel¹⁹, and differs from that for the native channel (ref. 13 and Fig. 4); it may reflect the presence of additional subunit species in the native channel^{30,31}. As for the reduction in maximum cAMP-induced current by Ca^{2+} -calmodulin, this probably results from a lower open probability of the channel, because the single-channel conductances were similar in the presence and absence of Ca^{2+} -calmodulin (data not shown). This feature is not observed in the native channel (Fig. 4).

METHODS. Procedures for transient transfection of the cloned rat olfactory channel cDNA into HEK293 cells followed ref. 19. Excised inside-out patch recordings were made 48–72 h after transfection using borosilicate glass electrodes with tip lumen diameters of $\sim 1 \mu\text{m}$. Zero Ca^{2+} solution contained 140 mM NaCl, 5 mM KCl, 1 mM Na-EGTA, 10 mM Na-HEPES, pH 7.4; the $50 \mu\text{M}$ Ca^{2+} solution had $704 \mu\text{M}$ CaCl_2 and 2 mM Na-NTA (nitrilotriacetic acid) instead of Na-EGTA to give the correct buffered Ca^{2+} concentration. The patch electrode was filled with 0 Ca^{2+} solution. In a and b, the effects of calmodulin and mastoparan were assayed after the membrane patch was perfused with the corresponding solution for at least 1 min. In the mastoparan experiment, mastoparan and calmodulin were added together after the current had recovered from an inhibition by Ca^{2+} -calmodulin. In c and d, the dose-response relation was measured first in zero Ca^{2+} , then after the patch had been perfused with test solution for 1 min.

the native channel (see later) and will not be considered here. The modulation described seems to involve a direct interaction between Ca^{2+} -calmodulin and the channel. The lack of ATP precludes the involvement of a Ca^{2+} -calmodulin-activated kinase. Ca^{2+} -calmodulin had a similar action when a hydrolysis-resistant cyclic nucleotide analogue (8-bromo cGMP) was used, arguing against the action of a Ca^{2+} -calmodulin-activated

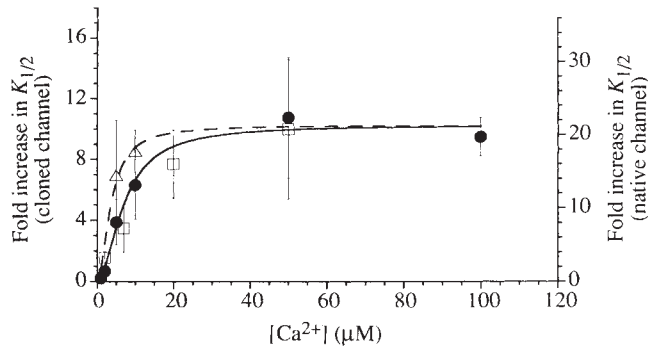


FIG. 2 Dependence of $K_{1/2}$ shift on Ca^{2+} concentration in the presence of calmodulin. Membrane voltage was -60 mV. Data from the cloned (circles and triangles) and native channels (squares) are shown. Cloned channel experiments were done with cGMP, native channel experiments with cAMP. Circles and squares, 235 nM calmodulin; triangles, 1 μM calmodulin. At each Ca^{2+} concentration a set of dose-response curves (like those in Figs 1c, right, and 4b) was constructed from each of several membrane patches, and the increase in $K_{1/2}$ due to Ca^{2+} -calmodulin measured and averaged. Data are averaged values, with error bars indicating standard deviations. Numbers of patches used and Ca^{2+} concentrations were as follows. Circles: 4 (1 μM Ca^{2+}), 4 (2 μM), 5 (5 μM), 7 (10 μM), 7 (50 μM) and 5 (100 μM); triangles: 5 (5 μM Ca^{2+}) and 3 (10 μM); squares: 4 (2 μM Ca^{2+}), 5 (7 μM), 4 (20 μM) and 6 (50 μM). Solutions were as for Fig. 1, except that they contained 2 mM Na-NTA and the following CaCl_2 concentrations: 20 μM (1 μM free Ca^{2+}), 40 μM (2 μM), 98 μM (5 μM), 135 μM (7 μM), 188 μM (10 μM), 346 μM (20 μM), 704 μM (50 μM) and 1.087 mM (100 μM). Curves were fitted with the Hill equation: solid curve has a half-maximal shift at 7.1 μM Ca^{2+} and a Hill coefficient of 1.8 ; dashed curve was constructed assuming the same Hill coefficient and maximum shift, but with a half-maximal shift at 3.7 μM Ca^{2+} . For the native channel, the ordinate scale (right) was adjusted so that the point at 50 μM Ca^{2+} fell on the solid curve for ready comparison with the cloned channel data.

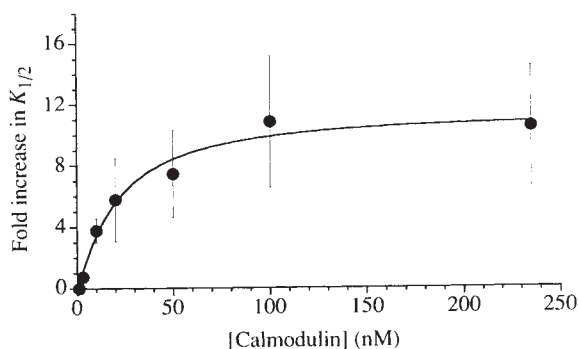


FIG. 3 Dependence of $K_{1/2}$ shift on calmodulin concentration for the cloned channel. Ca^{2+} concentration was 50 μM ; membrane voltage was -60 mV. At every calmodulin concentration a set of dose-response curves (like those in Fig. 1c, right) was constructed for each membrane patch using cGMP to measure the shift in $K_{1/2}$; the small $K_{1/2}$ shift due to 50 μM Ca^{2+} alone (without calmodulin; see Fig. 1c, left) has been subtracted. The dose-response curve in the presence of calmodulin was constructed only after steady-state inhibition had been reached. Data are averaged from several patches, with numbering 3 (1 nM calmodulin), 4 (3 nM), 5 (10 nM), 4 (20 nM), 5 (50 nM), 6 (100 nM) and 7 (235 nM), respectively. Curve is the Hill equation with a half-maximal shift at 21 nM calmodulin and a coefficient of 1.1 .

phosphodiesterase which might reduce the effective cyclic nucleotide concentration around the channel.

The dependence of the $K_{1/2}$ shift on Ca^{2+} concentration is shown in Fig. 2. In the presence of 235 nM calmodulin (filled circles), the fitted curve corresponds to a half-maximal shift at 7.1 μM Ca^{2+} and a Hill coefficient of 1.8 . The calmodulin concentration in olfactory cilia has been measured to be at least 1 μM , possibly much higher²⁰. At 1 μM calmodulin, extrapolation of the data (triangles) gives a half-maximal shift at ~ 3.7 μM Ca^{2+} . Figure 3 shows the dependence of $K_{1/2}$ shift on calmodulin concentration. The Ca^{2+} concentration was 50 μM , which should ligand most or all of the calmodulin²³, so the results roughly reflect the dependence of the shift on the concentration of the Ca^{2+} -calmodulin complex. The curve fitted to the points has a Hill coefficient of 1.1 and a half-maximal shift at 21 nM calmodulin. It could have been deduced from the Hill coefficient that one bound Ca^{2+} -calmodulin complex is sufficient to produce the shift even though the channel is an oligomer; but the situation is probably more complex, because $K_{1/2}$ is an operational parameter and does not reflect only cGMP binding to the channel.

Does this modulation occur in the native channel? A soluble factor involved in a similar Ca^{2+} modulation of the native olfactory channel in catfish was previously reported not to be mimicked by calmodulin¹⁶. We therefore made recordings from membrane patches excised from single, dissociated rat olfactory receptor neurons (Fig. 4 legend). In agreement with previous results¹³, the control $K_{1/2}$ of the native channel for cAMP was ~ 3 – 4 μM at -60 mV; 50 μM Ca^{2+} alone had only a slight effect (Fig. 4a), but in the presence of 235 nM calmodulin it shifted the $K_{1/2}$ by 19.4 -fold (Fig. 4b). At $+60$ mV, the shift was 21 -fold. The measured Ca^{2+} -dependence of this shift in cAMP $K_{1/2}$ was like that for the cloned channel (Fig. 2, squares). However, unlike the cloned channel, the maximum current through the native channel was not affected by Ca^{2+} -calmodulin. We also found a large modulation by Ca^{2+} -calmodulin in the native olfactory channel in newt and catfish (data not shown), so the discrepancy of our results with ref. 16 is not due to a species difference. Because in most of their experiments Kramer *et al.*¹⁶ excised the entire dendrite plus cilia and not just a small membrane patch from an olfactory receptor neuron for recording, exogenously applied calmodulin might not have gained full access to the cilia, accounting for their negative result. Also, their choice of calmidazolium as a test for calmodulin involvement was complicated by its inhibitory effect on the channel itself.

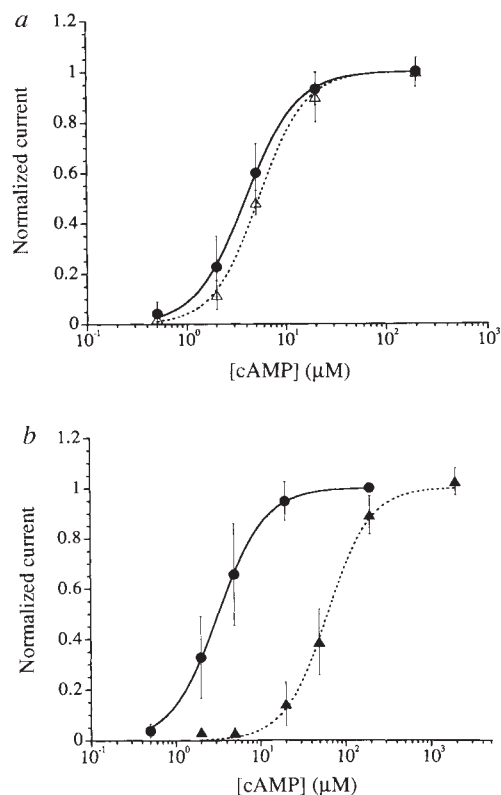
We have found that Ca^{2+} -calmodulin has a strong effect on the apparent ligand affinity of the olfactory cyclic nucleotide-activated channel. A Ca^{2+} -calmodulin effect has also been reported on adenylate cyclase and phosphodiesterase in olfactory cilia^{20,24,25}. But as adenylate cyclase activity increases with Ca^{2+} -calmodulin^{20,24}, it is unlikely to contribute to adaptation. The overall activity of the phosphodiesterase in cilia varies only threefold in low and high Ca^{2+} concentrations in the presence of calmodulin²⁵. By comparison, the ligand affinity of the channel is altered over 20 -fold by Ca^{2+} -calmodulin, or 200 – 400 -fold in terms of channel gain as calculated from the Hill coefficient. This channel modulation is therefore likely to be important for Ca^{2+} -mediated adaptation. Further quantitative predictions cannot be made as the free- Ca^{2+} concentration in the cilia of an intact receptor neuron during odorant stimulation is not known. But on the basis of the size of the Ca^{2+} -activated chloride current^{26,27} induced by odorants in intact neurons, the free- Ca^{2+} concentration could reach micromolar levels (Fig. 1 of ref. 28), which should activate the Ca^{2+} -calmodulin modulation of the channel at least partially.

The cGMP-activated channel mediating visual transduction in retinal rods is also affected by Ca^{2+} -calmodulin²². In this case Ca^{2+} -calmodulin was proposed to bind not to the 63K rod channel protein but to a 240K cytoskeleton-like protein tightly associated with it; modulation is then effected by allosteric inter-

FIG. 4 Effect of Ca^{2+} -calmodulin on the activation of the native rat olfactory channel by cAMP. Current was measured at -60 mV and normalized with respect to the value at $200 \mu\text{M}$ cAMP in zero- Ca^{2+} solution; procedure was as for Fig. 1c and d and data were averaged. a, comparison for 0 Ca^{2+} (circles) and $50 \mu\text{M}$ Ca^{2+} (triangles) conditions. Averaged data from 3 patches, with saturating currents at -60 mV ranging 47 – 224 pA. Solid curve has $K_{1/2} = 4.0 \mu\text{M}$, $n = 1.7$; dashed curve has $K_{1/2} = 5.4 \mu\text{M}$, $n = 1.9$. b, Comparison for 0 Ca^{2+} (circles) and $50 \mu\text{M}$ Ca^{2+} with 235 nM calmodulin (triangles). Averaged data from 6 patches (except for the data points at $2 \mu\text{M}$ and 2 mM cAMP with calmodulin, which were derived from 2 and 4 experiments, respectively), with saturating currents at -60 mV, range 8 – 193 pA. Solid curve has $K_{1/2} = 3.2 \mu\text{M}$, $n = 1.6$; dashed curve has $K_{1/2} = 63 \mu\text{M}$, $n = 1.7$, giving a shift of 19.4 -fold.

METHODS. Adult female Sprague-Dawley rats were killed by CO_2 asphyxiation. The olfactory turbinates were removed from the sagittally hemisected head and placed in cold Tyrode's solution (140 mM NaCl, 5 mM KCl, 2 mM CaCl_2 , 1 mM MgCl_2 , 10 mM glucose, 10 mM Na-HEPES, pH 7.4). Strips of olfactory epithelium were separated from the supporting cartilage and incubated in a dissociation solution (140 mM NaCl, 5 mM KCl, 10 mM glucose, 10 mM Na-HEPES, pH 7.4, with or without 0.02% trypsin) for 30 min at 37°C . The dissociation was stopped by washing the tissue with Tyrode's solution three times. After gentle trituration with a fire-polished Pasteur pipette, supernatant containing isolated cells was transferred onto a glass coverslip precoated with concanavalin A and allowed to settle so that the cells became immobilized. The coverslip was transferred to the recording chamber and perfused with Tyrode's solution. Olfactory receptor neurons were recognized by their anatomical features^{9,13}. Inside-out membrane patches were excised from the dendritic knob region of the cells using patch electrodes with tip lumen diameters of $\sim 0.2 \mu\text{m}$ and resistances of 15 – $30 \text{ M}\Omega$. The electrode contained zero- Ca^{2+} solution; solutions were as for Fig. 1.

action between the 63K and 240K proteins²². Our results, however, indicate that the action of Ca^{2+} -calmodulin on the olfactory channel is direct because expression of the cloned olfactory channel protein alone is enough to produce the Ca^{2+} -calmodulin effect. It might be argued that the 293 cells we used for channel expression happened to contain the 240K protein found in rods, but this is unlikely because the cloned 63K rod channel protein when expressed in the same cells failed to show any Ca^{2+} -calmodulin effect (data not shown). The mechanism by which Ca^{2+} -calmodulin modulates the rod channel remains to be determined. □



Ventralizing signal determined by protease activation in *Drosophila* embryogenesis

Cynthia L. Smith & Robert DeLotto

Molecular Biology Department,
Sloan-Kettering Institute for Cancer Research,
1275 York Avenue and Department of Cell Biology,
Cornell University Graduate School of Medical Sciences,
1300 York Avenue, New York, New York 10021, USA

SPECIFICATION of dorsal-ventral cell fate during *Drosophila* embryogenesis is mediated by a signal transduction pathway^{1–4}. Asymmetry of cell fates arises through the spatially restricted production of a ligand in an extracellular compartment called the perivitelline space⁵. The *snake* and *easter* genes are required for the production of the ligand¹⁷ and they encode the proenzyme form of secreted extracellular serine proteases^{6,7}. We have examined the effect of producing a preactivated form of the snake protease on the generation of dorsal-ventral polarity. SP6 RNA microinjection experiments reveal that different cell fates acquired at cellular blastoderm can be specified by the amount and spatial distribution of activated *snake* protein. Our results support a protease cascade model in which localized activation of uniformly distributed protease proenzymes leads to the spatially restricted production of ligand in the perivitelline space on the ventral side of the embryo.

To test whether localized activation of *snake* is directly involved in the generation of the ventralizing signal, a cleaved and activated form of the snake protease catalytic chain was introduced into embryos. The wild-type *snake* protease is secreted as an inactive zymogen which requires proteolytic cleavage for activation. A variant *snake* SP6 RNA transcript called

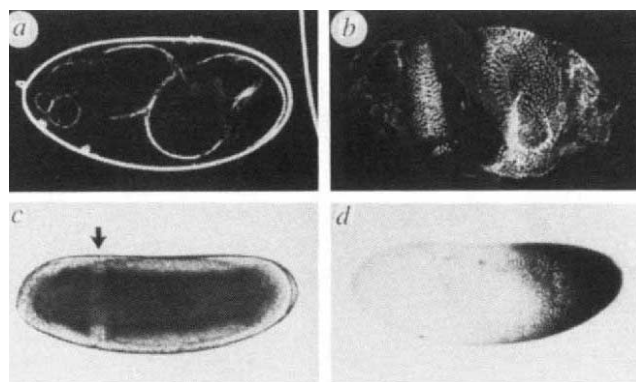
Received 6 December 1993; accepted 23 February 1994.

- Firestein, S. & Werblin, F. *Science* **244**, 79–82 (1989).
- Kurahashi, T. *J. Physiol., Lond.* **419**, 177–192 (1989).
- Lowe, G. & Gold, G. H. *J. Physiol., Lond.* **442**, 147–168 (1991).
- Pace, U., Hanski, E., Salomon, Y. & Lancet, D. *Nature* **316**, 255–258 (1985).
- Sklar, P. B., Anholt, R. R. H. & Snyder, S. H. *J. Biol. Chem.* **261**, 15538–15543 (1986).
- Jones, D. T. & Reed, R. R. *Science* **244**, 790–795 (1989).
- Bakalyar, H. A. & Reed, R. R. *Science* **250**, 1403–1406 (1990).
- Boekhoff, I., Tareilus, E., Strotmann, J. & Breer, H. *EMBO J.* **9**, 2453–2458 (1990).
- Nakamura, T. & Gold, G. H. *Nature* **325**, 442–444 (1987).
- Kurahashi, T. *J. Physiol., Lond.* **430**, 355–371 (1990).
- Firestein, S., Darrow, B. & Shepherd, G. M. *Neuron* **6**, 825–835 (1991).
- Frings, S. & Lindemann, B. *J. gen. Physiol.* **97**, 1–16 (1991).
- Frings, S., Lynch, J. W. & Lindemann, B. *J. gen. Physiol.* **100**, 45–67 (1992).
- Lowe, G. & Gold, G. H. *J. Physiol., Lond.* **462**, 175–196 (1993).
- Kurahashi, T. & Shibuya, T. *Brain Res.* **515**, 261–268 (1990).
- Kramer, R. H. & Siegelbaum, S. A. *Neuron* **9**, 897–906 (1992).
- Zufall, F., Shepherd, G. M. & Firestein, S. *Proc. R. Soc. Lond.* **B246**, 225–230 (1991).
- Kleene, S. J. *J. Membr. Biol.* **131**, 237–243 (1993).
- Dhallan, R. S., Yau, K.-W., Schrader, K. A. & Reed, R. R. *Nature* **347**, 184–187 (1990).
- Anholt, R. R. H. & Rivers, A. M. *Biochemistry* **29**, 4049–4054 (1990).
- Chen, T.-Y. *et al.* *Nature* **362**, 764–767 (1993).
- Hsu, Y.-T. & Molday, R. S. *Nature* **361**, 76–79 (1992).
- Klee, C. B. & Vanaman, T. C. *Adv. Prot. Chem.* **35**, 213–321 (1982).
- Jaworsky, D. E., Matsuzaki, O., Borisy-Rudin, F. & Ronnett, G. V. *J. Neurosci.* (in the press).
- Borisy, F. F. *et al.* *J. Neurosci.* **12**, 915–923 (1992).
- Kurahashi, T. & Yau, K.-W. *Nature* **363**, 71–74 (1993).
- Lowe, G. & Gold, G. H. *Nature* **366**, 283–286 (1993).
- Kleene, S. J. & Gesteland, R. C. *J. Neurosci.* **11**, 3624–3629 (1991).
- Kleene, S. J. *Br. J. Pharmac.* **111**, 469–472 (1994).
- Bradley, J., Uezono, Y., Davidson, N. & Lester, H. Soc. Neurosci. Abstr. **18**, 596 (1992).
- Liman, E. R. & Buck, L. B. *Biophys. J.* **66**, A350 (1994).

ACKNOWLEDGEMENTS. We thank R. Huganir, Y. Koutalos, T. Kurahashi, R. S. Molday, J. Nathans, R. R. Reed, G. Yellen and D. T. Yue for discussion, and S. J. Kleene for sharing his data on the calmidazolium action²⁹. This work was partly supported by a grant from the US National Eye Institute.

FIG. 1 Lateralization and ventralization of embryos from *snk*⁻ females by microinjection of *snkΔ_{nc}* RNA. a, An embryo from a *snk*²²⁹/*snk*²²⁹ female. b, An embryo injected posteriorly with *snkΔ_{nc}* RNA (200 μg ml⁻¹). c, Gastrulation pattern of an embryo injected centrally with RNA at 100 μg ml⁻¹. d, An embryo injected posteriorly with *snkΔ_{nc}* RNA at 2 mg ml⁻¹ and stained with anti-twist antibodies¹²⁻¹⁴.

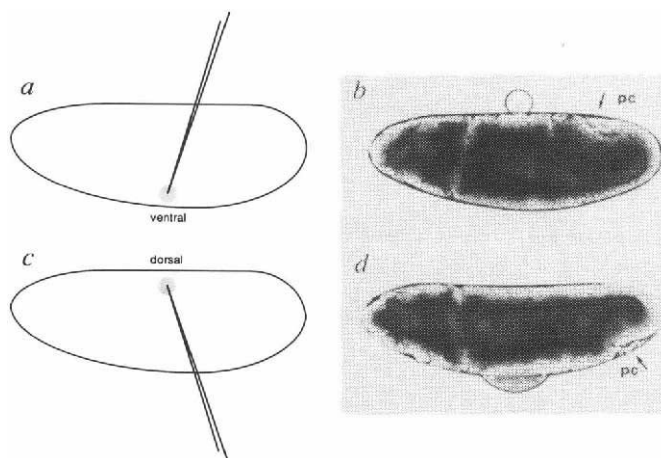
METHODS. *snkΔ_{nc}* construction. The 5' untranslated leader and signal peptide of *easter* complementary DNA was fused in-frame to the catalytic chain of snake. Full-length cDNA was made from 0-3-h-old embryos. PCR was used to isolate a DNA fragment containing the 5' untranslated region and signal peptide of *easter*. PCR primers were 5'-CTAGAATTCAGATCTGAGTTGTGCGCAAGCTCAGCTTCGGC-3' and 5'-TCAGGTACCGCCGGCCGATGATTCGCCAAATGCCAA-3'. *Eco*RI and *Kpn*I sites were used to clone the DNA fragment into pGEM4 (Promega). Site-directed mutagenesis¹⁵ using primer 5'-CGTCCTGCTTCCCTGTGGCA-3' was used to alter cysteine 323 of the snake catalytic chain to a serine. The altered catalytic chain was prepared by PCR with 5'-GACGG ATCCCTAGTGTGCTTGAAGGCAT-3' and 5'-CGATCGGCCGCGCATAGTCGGT GGAACCCCACT-3'. This PCR fragment was cloned into the first plasmid using *Bam*HI and *Eag*I. *snkΔ_{nc}* (Ser to Ala) was made similarly using snake S381A as a PCR template¹⁰. SP6 RNA was prepared and stage 2 embryos were injected centrally from the posterior unless specified and at 65% egg length in c with



2% egg volume of *snkΔ_{nc}* RNA in water. Cuticles were prepared as described¹⁶. Anti-twist staining of embryos injected with *snkΔ_{nc}* at 2 mg ml⁻¹ at 10% egg length. Embryos were fixed in 4% paraformaldehyde (Fluka) and stained with polyclonal anti-twist antisera (provided by S. Roth) using horseradish peroxidase-conjugated goat-anti-rabbit secondary antibody.

FIG. 2 Polarization of the dorsal-ventral axis by *snkΔ_{nc}*. a, b, *snkΔ_{nc}* SP6 RNA injected into the yolk on the ventral side of the egg, caused the embryo to gastrulate with normal dorsal-ventral polarity and asymmetry (note the position of pole cells and angle of the cephalic furrow). c, d, RNA injected into the yolk on the dorsal side of the egg caused the embryo to gastrulate with normal asymmetry but with inverted polarity with respect to the egg shape. PC, Pole cells.

METHODS. *snkΔ_{nc}* RNA at a concentration of 200 μg ml⁻¹ was microinjected into the cortical ooplasm on the ventral side of the egg by traversing the egg diameter with the injection needle. The embryos were allowed to gastrulate and photographed in brightfield. Cuticle preparations of asymmetrically injected embryos showed ventral structures on the side of the egg where RNA was deposited.



snkΔ_{nc}, encodes the snake catalytic chain fused in-frame to a signal peptide. This transcript is designed to direct secretion of the 'activated' snake catalytic chain into the perivitelline space when microinjected into embryos.

In vitro transcribed *snkΔ_{nc}* RNA was microinjected at different concentrations into embryos derived from *snk*²²⁹/*snk*²²⁹ mothers. Uninjected embryos are completely dorsalized. Embryos were scored for the production of specific ventrolateral pattern elements such as filzkörper, lateral hairs and ventral denticles and analysed by their gastrulation pattern. When *snkΔ_{nc}* RNA was injected in a central position, embryos developed ventrolateral cuticular pattern elements in rings typical of lateralized phenotypes (Fig. 1b) and the pattern of gastrulation was lateralized, lacking the most extreme dorsal or ventral cuticular features and exhibiting an expansion of intermediate cuticular features (Fig. 1c). Table 1a quantifies the formation of ventrolateral pattern elements as a function of the *snkΔ_{nc}* RNA concentration. Although RNA at 50 μg ml⁻¹ produces primarily dorsalized embryos and filzkörper, RNA at 200 μg ml⁻¹ produces ventral denticle bands in 29% of injected embryos. Higher concentrations of RNA (2 mg ml⁻¹) produced lateralized embryos with fewer hard cuticle elements.

To test whether very high concentrations of *snkΔ_{nc}* (2 mg ml⁻¹) shift cell fates to that of extremely ventral mesoderm, antibody staining for twist protein was used. The twist protein is specifically expressed in blastoderm nuclei of mesodermal primordia⁸. Dorsalized embryos from *snk*²²⁹/*snk*²²⁹ females do not express twist. *snkΔ_{nc}* RNA (2 mg ml⁻¹) was injected pos-

teriorly and embryos were stained with anti-twist antibodies. Nuclei of all cells along the dorsal-ventral axis on the posterior side of the embryo stained darkly for twist protein (Fig. 1d). These results collectively show that the relative cell fate along the dorsal-ventral axis changes as a function of the concentration of *snkΔ_{nc}* RNA. Cuticle preparations and antibody staining also reveal that the phenotypic effect of *snkΔ_{nc}* RNA injection is near the site of injection. *snkΔ_{nc}* (S to A) transcripts, with the protease active-site serine altered to alanine, fail to lateralize (Table 1a). This argues that the injection phenotype requires snake protease activity.

If activation of snake proenzyme is rate limiting for the production of ligand, microinjection of *snkΔ_{nc}* might have a dominant effect in wild-type embryos. To test this, *snkΔ_{nc}* RNA was injected into wild-type embryos. Of 90 embryos injected, all exhibited cuticle pattern defects consistent with lateralization or ventralization. In most cases, cuticles corresponded to the V1 (ventralized) classification of mutant phenotypes exhibiting partial ventral epidermis and mesodermal derivatives⁹.

To determine which genes are required for the conversion of the snake proenzyme to active form and which genes are required to transmit the snake signal, *snkΔ_{nc}* RNA was microinjected into embryos from females homozygous for strong mutant alleles of various dorsal-group genes (Table 1b). Embryos were injected into the centre at 50% egg length and scored by gastrulation patterns. Gastrulation was lateralized or ventralized in embryos from females homozygous for mutations in the genes *nudel*, *pipe*, *windbeutel* and *gastrulation defective*. The pattern of gastrulation

TABLE 1 Phenotypic effects of the *snkΔn_e* RNA microinjection

(a) Ventrolateral pattern elements versus RNA concentration							
[RNA]	NO	NC*(%)	DO(%)	FK(%)	VD(%)	Gastrulation	Twist expression
<i>snkΔn_e</i> (S to A)							
200 μg ml ⁻¹	84	100	0	0	0	dorsalized	—
<i>snkΔn_e</i>							
0	95	100	0	0	0	dorsalized	—
50 μg ml ⁻¹	96	64	17	17	2	dorsalized/lateralized	—
200 μg ml ⁻¹	83	59	5	7	29	lateralized	+
2 mg ml ⁻¹	79	87	0	3	10	lateralized	++
(b) Gastrulation phenotype of <i>snkΔn_e</i> in dorsitized mutant backgrounds							
Maternal genotype	NO	Gastrulation phenotype					
Wild type	90	lateralized/ventralized					
<i>nud⁰⁹³/nud⁰⁹³</i>	73	lateralized/ventralized					
<i>wbl^{RP54}/wbl^{RP54}</i>	36	lateralized/ventralized					
<i>gd⁸/gd⁸</i>	35	lateralized/ventralized					
<i>pip⁶⁶⁴/pip⁶⁶⁴</i>	58	lateralized/ventralized					
<i>ea¹/ea¹</i>	44	dorsitized					
<i>spz¹⁹⁷/spz¹⁹⁷</i>	35	dorsitized					

a, *snkΔn_e* RNA was microinjected in a posterior central position in stage 2 embryos at the indicated concentration and cuticles were scored after 36 h as per cent. NC, Undifferentiated or poorly differentiated cuticle; DO, dorsitized; FK, exhibiting filzkörper material; VD, exhibiting ventral denticles. NO, Total number of embryos injected. For analysis of twist expression, injected embryos were stained with anti-twist polyclonal antisera: (—) no twist expression; (+) moderate twist expression; (++) high levels of twist expression. b, RNA concentration was 250 μg ml⁻¹ and injected into stage 2 embryos as a bolus in a central position at 65% egg length in embryos derived from females of the indicated (strong) mutant allele. Embryos were allowed to gastrulate at 25 °C and were scored as dorsitized if all embryos exhibited a pattern of gastrulation typical of uninjected recipients or lateralized/ventralized if either the gastrulation pattern was typical of lateralization or if a broad asymmetric ectopic ventral furrow began to form near the site of injection.

* In this experiment uninjected *snk²²⁹/snk²²⁹* control embryos were poorly differentiated and therefore classified as NC, but were unambiguously dorsitized by virtue of their gastrulation pattern.

was dorsitized in embryos from females homozygous for mutations of *easter* and *spatzle*. Cuticles made from injected *pipe*, *windbeutel* and *gastrulation defective* embryos exhibited ventrolateral pattern elements including ventral denticles. Therefore, the dominant effect of *snkΔn_e* injection works through the products of the *easter* and *spatzle* genes and does not require the gene products of *nudel*, *pipe*, *windbeutel* and *gastrulation defective*. These four gene products are therefore required for snake to become activated.

Asymmetric injection of *snkΔn_e* RNA can polarize the gastrulation pattern. Wild-type embryos have a stereotypical asymmetry in their gastrulation pattern, whereas completely dorsitized embryos lack dorsal-ventral asymmetry. Wild-type *snake* RNA transcripts do not polarize the pattern of gastrulation as a function of the site of injection¹⁰. To determine whether *snkΔn_e* RNA can polarize *snake* embryos, *snkΔn_e* transcripts were microin-

jected into the cytoplasm either on the dorsal or the ventral side of the egg and the gastrulation pattern was analysed.

When RNA was microinjected on the ventral side of the egg (Fig. 2a), the gastrulation pattern exhibited normal polarity (Fig. 2b). Pole cells migrated toward the dorsal side (arrow) and the cephalic furrow formed with its correct angle with respect to the egg shape. When RNA was introduced into the dorsal cytoplasm (Fig. 2c), the polarity of the gastrulation was reversed (Fig. 2d). Pole cells migrated toward the ventral side of the egg and the angle of the cephalic furrow was reversed relative to wild-type embryos.

Data presented here indicate that the level of 'activated' snake protease can define cell fates along the dorsal-ventral axis in a concentration-dependent manner. A lateralized cuticle and gastrulation phenotype is observed when RNA is injected into the centre of the embryo. RNA placed in an asymmetric position in the cytoplasm produces a ventralized cuticle and gastrulation phenotype. A possibility is that wild-type embryos generate a snake activation gradient with a high point on the ventral side and a low point on the dorsal side.

Embryos produced by females lacking *easter* function are lateralized when injected with *eaΔN* RNA which encodes a proenzyme polypeptide deletion¹¹. Observations for *snkΔn_e* and *eaΔN* injection are consistent with both proteases being in a sequential protease activation pathway. The combined results strongly support the protease cascade model for the interaction of snake and *easter* with other components of the dorsal-ventral pathway⁶. A cascade model based on the epistasis data is schematically illustrated in Fig. 3. The cascade model accounts for the presence of the rescue activities as cytoplasm and messenger RNA are in excess and exhibit no apparent spatial asymmetry in the embryo. The ventralizing signal would arise by local activation of uniformly distributed protease zymogens. The relative amount of the proteases that are in an activated state is probably very small compared to the amount of zymogen and the process might occur rapidly. At least one upstream component of the pathway that initiates the protease activation cascade and orients polarity must be symmetrically distributed. □

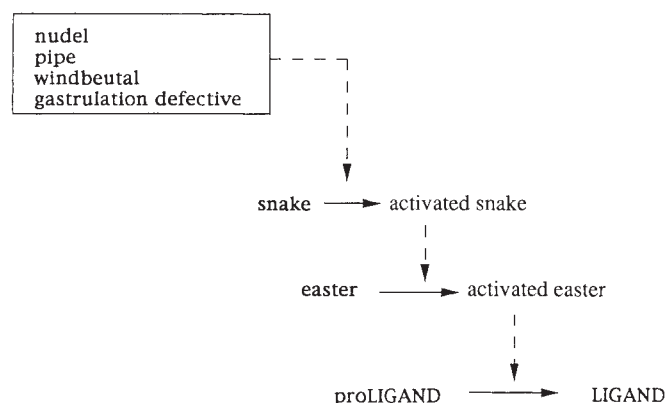


FIG. 3 A protease cascade model. Activation of the zymogen form of *easter* requires the activation of the zymogen form of *snake* upstream in the cascade. *nudel*, *pipe*, *windbeutel* and *gastrulation defective* function (boxed) are required upstream of *snake* for its activation. Relative order of these four components is unknown. *Easter* zymogen activation requires *snake* activity¹¹. Dotted lines indicate relative order of components as determined by epistasis but do not necessarily imply direct interactions.

Received 1 October 1993; accepted 16 February 1994.

1. Govind, S. & Steward, R. *Trends Genet.* **7**, 119–125 (1991).

2. Anderson, K. *Trends Genet.* **3**, 91–97 (1987).
3. Rushlow, C. & Arora, K. *Semin. Cell Biol.* **1**, 137–149 (1990).
4. St Johnson, D. & Nusslein-Volhard, C. *Cell* **68**, 201–219 (1992).
5. Stein, D., Roth, S., Vogelsang, E. & Nusslein-Volhard, C. *Cell* **65**, 725–735 (1991).
6. DeLotto, R. & Spierer, P. *Nature* **323**, 688–692 (1986).
7. Chasen, R. & Anderson, K. V. *Cell* **56**, 391–400 (1989).
8. Leptin, M. & Grunewald, B. *Development* **110**, 73–84 (1990).
9. Roth, S., Hiromi, Y., Godt, D. & Nusslein-Volhard, C. *Development* **112**, 371–388 (1991).
10. Smith, C. L. et al. *Genetics* **136**, 1355–1365 (1994).
11. Chasen, R., Jin, Y. & Anderson, K. *Development* **115**, 607–616 (1992).
12. Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
13. Santamaria, P. & Nusslein-Volhard, C. *EMBO J.* **2**, 1695–1699 (1983).
14. Santamaria, P. in *Drosophila, a Practical Approach* (ed. Roberts, D. B.) 159–174 (IRL, Oxford, Washington DC, 1986).
15. Kunkel, T., Roberts, J. & Zakour, R. *Meth. Enzym.* **154**, 367–382 (1987).
16. Wieschaus, E. & Nusslein-Volhard, C. in *Drosophila, a Practical Approach* (ed. Roberts, D. B.) 199–228 (IRL, Oxford, Washington DC, 1986).
17. Stein, D. & Nusslein-Volhard, C. *Cell* **68**, 429–440 (1992).

ACKNOWLEDGEMENTS. We thank Y. DeLotto for assistance in microinjections, J. Heyer for staining methods, T. Schupbach for providing stocks and M. Osley and D. Stein for comments and criticism of earlier versions of the manuscript.

An essential role for HLA-DM in antigen presentation by class II major histocompatibility molecules

Phillip Morris*, Jeffrey Shaman*, Michelle Attaya†, Miguel Amaya*, Steven Goodman*, Carolyn Bergman*, John J. Monaco†‡ & Elizabeth Mellins*§

* Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

† Department of Microbiology and Immunology, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298-0678, USA

In antigen-presenting cells, class II molecules of the major histocompatibility complex (MHC) bind peptides derived from endocytosed proteins¹. In certain B-lymphoblastoid cell mutants, MHC class II molecule-peptide complex formation is impaired, resulting in deficient antigen-presenting function². MHC deletion mutants with this defect map the responsible gene(s) to the class II region of the MHC^{3–5}. Here we report that multiple independent mutants with the class II presentation defect harbour lesions in *HLA-DMB*, an MHC-linked gene encoding a class II-like β -chain^{6,7}. Expression of *DMB* complementary DNA in mutants lacking *DMB* messenger RNA restores the wild-type phenotype. These results establish HLA-DM as a critical regulatory molecule in class II-restricted antigen presentation and suggest that it functions at an intracellular site to promote class II molecule-peptide association.

Sixteen B-lymphoblastoid cell mutants were derived from an *HLA-DR/DQ* hemizygous progenitor, designated 8.1.6, by several replicate immunoselection experiments using a DR3-specific monoclonal antibody, 16.23 (Fig. 1a, and refs 2, 8). The phenotype of these mutants has been described². Their key feature is a defect in the presentation of intact protein antigen, but not of exogenously supplied peptides, to class II-restricted T cells. In addition, class II molecules expressed by the mutants lack characteristics of the mature, peptide-loaded molecules of the progenitor cell, such as expression of certain antibody epitopes (for example, antibody 16.23) and stability in SDS detergent solutions, although the molecules are expressed in normal amounts at the cell surface. Analysis of HLA-DR-peptide complexes from the mutants reveals that 60–70% of DR

molecules are associated with a nested set of peptides from residues 80–103 of the invariant chain (Ii)^{9–11}, a class II molecule chaperone. Similar class II molecule-Ii peptide complexes have been found in wild-type cells^{1,12,13}, implying that this complex may be a biosynthetic intermediate that accumulates in the mutants.

The derivation of the mutants from a progenitor that is hemizygous for the class II region of the MHC (8.1.6; Fig. 1a) suggested that the responsible gene(s) mapped to this region. Some MHC deletion mutants, including the 8.1.6-derived mutant 5.2.4 (Fig. 1a), were found to share the class II presentation defect^{3–5}, and analysis with such mutants located the relevant gene(s) within the centromeric ~230 kilobases (kb) of the region deleted in progenitor 8.1.6. To refine the mapping of this gene, we established the centromeric breakpoint of the 8.1.6 deletion. Southern blot analyses of mutant 5.2.4 indicated the presence of *RING-3* (ref. 14) and the absence of *LMP-2* (P.M. and J.S., unpublished data) on the *DR1* haplotype of 8.1.6 and placed the breakpoint between these two loci (Fig. 1b); further mapping studies indicated that the breakpoint lay within the *DMB* gene (Fig. 1b legend).

The disruption of one copy of the *DMB* gene in progenitor 8.1.6 suggested that mutation of the remaining *DMB* gene (*DR3* haplotype) might be responsible for the class II presentation defect in 8.1.6-derived mutants. Southern blot analysis showed no alteration of the *DR3*-linked *DMB* gene in any mutant except 5.2.4 (data not shown). However, northern blot analysis of *DMB* transcripts from the mutants revealed a striking number of abnormalities (Fig. 2). No *DMB* message is detectable in mutant 5.2.4, indicating that the truncated *DMB* gene of the *DR1* haplotype in 8.1.6 (and its derivatives) does not produce stable transcripts. Northern analysis of *DMB* mRNA from the other mutants thus assesses the transcripts generated by the *DR3*-linked *DMB* gene. In 11 mutants, the *DMB* transcript is of normal size (1.4 kb), but is present in vastly reduced amounts. In three other mutants, the *DMB* transcript is moderately less abundant and is shorter (~1.3 kb). In one mutant, 7.19.6, the *DMB* RNA is only 1.1 kb long (Fig. 2). Polymerase chain reaction (PCR) amplification of *DMB* cDNA from all mutants with short transcripts indicates that each contains a mutation in the *DMB* gene rather than truncation at the 3' end of the gene (J.S., data not shown). These results implicate *DMB* as the gene responsible for the mutant phenotype.

Only one 16.23-selected mutant, 7.12.6, expresses a *DMB* transcript of normal length and abundance (Fig. 2). This mutant also has a unique phenotype. The reduction in binding of the 16.23 antibody to mutant 7.12.6 is less severe (compare with mutant 9.10.3; Fig. 3a), and several T-cell lines are moderately reactive to 7.12.6 cells as antigen-presenting cells (APC), although certain T-cell clones are unresponsive (Fig. 3b). These findings raised the possibility that the 7.12.6 defect was in some gene other than *DMB*. However, sequencing of 7.12.6 *DMB* cDNA revealed a single G to A change, which results in a tyrosine for cysteine substitution (J.S., data not shown). The mutated residue at position 79 corresponds to a conserved cysteine that participates in an intrachain disulphide bond in MHC class I and class II peptide-binding domains⁵.

The data from all 17 mutants thus strongly implicate mutation of *DMB* as the basis of the mutant phenotype. We therefore introduced *DMB* cDNA into several mutants and evaluated the transfectants, 9.5.3-*DMB*, 7.9.6-*DMB* and 7.2.6-*DMB*, for restoration of the wild-type phenotype. The *DR3* molecules expressed by the transfectants regained characteristics of *DR3* molecules from wild-type cells: they expressed the 16.23 antibody epitope (Fig. 4a) and showed increased stability in SDS-PAGE (Fig. 4b). Further, the *DMB* transfectants presented native antigens to class II-restricted T cells (Fig. 4c) and stimulated alloreactive DR- and DP-specific T cells, which do not recognize the mutants (Fig. 4c). Transfection of the murine *DMB* homologue, *Mb*, also increased binding of the 16.23 antibody

‡ Present address: Department of Molecular Genetics, University of Cincinnati, Cincinnati, Ohio 45267-0524, USA.

§ To whom correspondence should be addressed.

(7.2.6-Mb; Fig. 4a), whereas control cells with the pcDNA/neo vector (7.9.6-C, 7.2.6-C) maintained the mutant phenotype (Fig. 4a).

We have shown that a defect in antigen presentation by conventional HLA class II molecules results from mutation of *HLA-DMB*. The sequence of the *HLA-DMB* gene suggests that it encodes a class II-like β -chain; the linked *HLA-DMA* locus encodes a class II-like α -chain, implying that the products of these genes may associate to form a heterodimer⁷. The proposed

DM molecule and its murine homologue Ma/Mb show limited sequence similarity in the membrane-distal domain to the corresponding domains of other class II molecules (14–25%)^{6,7}. Indeed, analysis of the structure of the putative HLA-DM molecule by sequence alignment with HLA-DR1 predicts with only moderate confidence that HLA-DM folds to form a peptide-binding cleft, and alignment of DM β with other class II β -chains places several insertions and deletions in the β 1 helix (refs 6, 7, 15, and J. Brown, personal communication). This

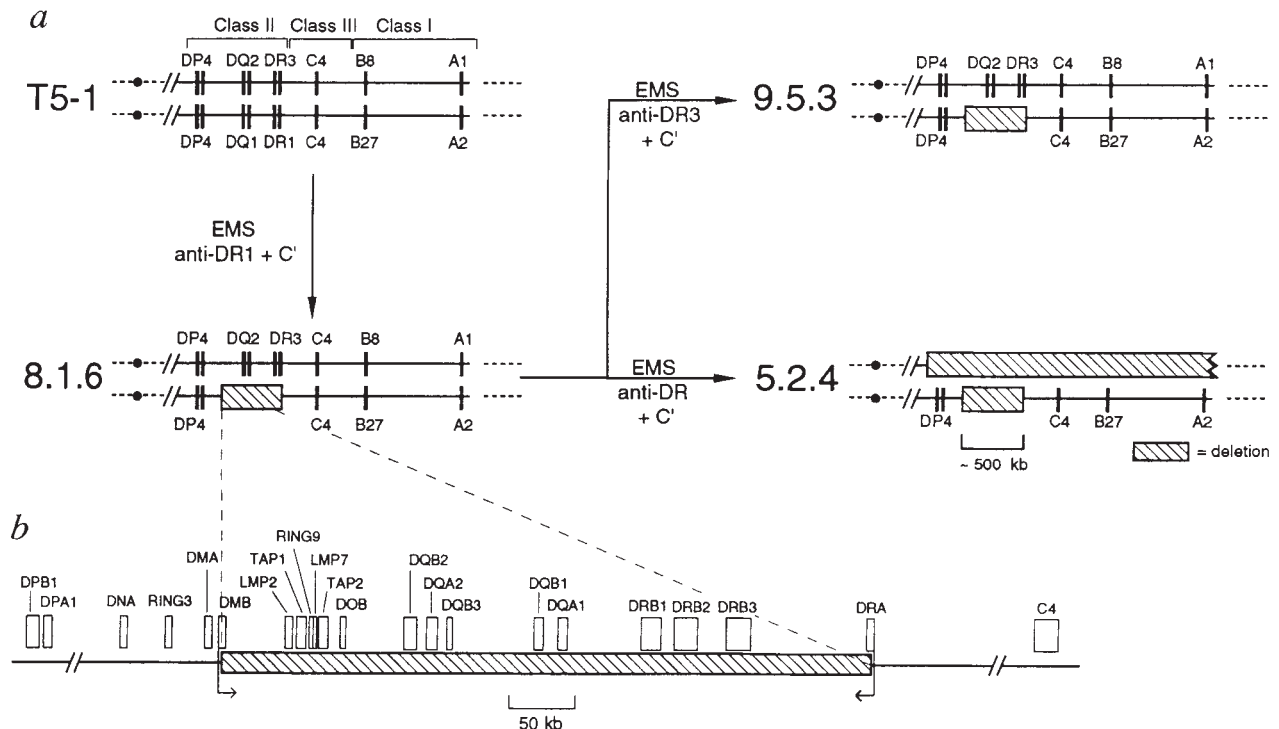


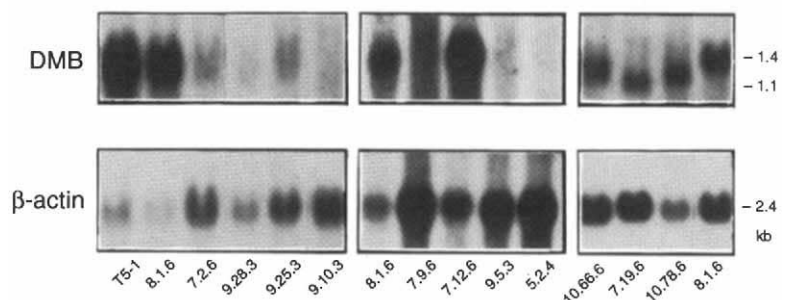
FIG. 1 Derivation of cell lines 8.1.6, 9.5.3 and 5.2.4 from B-lymphoblastoid cell line T5-1 and mapping of deletion breakpoints in 8.1.6. *a*, Immunoselection of 8.1.6 from T5-1 and subsequent immunoselections of 9.5.3 and 5.2.4 from 8.1.6. Cell lines are designated by extended MHC haplotypes. The vertical bars indicate the approximate locations of genetic loci (for example, *DRA* and *DRB*), which encode expressed proteins. Additional genes located in this region of the MHC (ref. 22) are not shown. *b*, Genetic map of the relevant portion of the MHC class II region (adapted from ref. 22, with permission) with delineation of the 8.1.6 deletion. Transcriptional orientations of *DMB* and *DRA* genes are indicated by arrows.

METHODS. Cell lines in *a* (8.1.6, 9.5.3, 5.2.4) were derived by ethyl methane sulphonate (EMS) mutagenesis of indicated parental lines and selection with antibody and complement (C') (selecting antibodies: anti-DR1 antiserum, anti-DR3 monoclonal antibody 16.23 (ref. 23), or

monomorphic anti-DR monoclonal antibody VI.15 (ref. 24)). These immunoselections have been described in detail elsewhere (refs 25, 2, 3, respectively). The centromeric deletion breakpoint within the *DMB* gene of 8.1.6 was identified by Southern blot analysis of *P*stI-digested DNA, which reveals a *DMB* restriction-fragment length polymorphism (13-kb fragment) in 8.1.6 and its derivatives, but not in T5-1 (P.M. and J.S., data not shown). This fragment hybridizes with a 200-bp probe from the 5'-untranslated region of *DMB*, but does not hybridize with a 185-bp probe from the 3'-untranslated region of *DMB* (primers used to generate the *DMB* 5' and 3' probes were a gift from D. Chaplin, Washington University, St Louis, MO). The telomeric breakpoint of the 8.1.6 deletion lies in the first exon of the *DRA* gene; as a result, a segment of *DRA* is also found on the 13-kb *P*stI fragment in 8.1.6 and its derivatives (P.M., data not shown).

FIG. 2 Northern blot hybridization of RNA samples from B-lymphoblastoid cell lines T5-1, 8.1.6 and class II-presentation-defective mutants (7.2.6, 9.28.3, 9.25.3, 9.10.3, 7.9.6, 7.12.6, 9.5.3, 5.2.4, 10.66.6, 7.19.6, 10.78.6) with *DMB* cDNA (upper) and β -actin (lower) probes. A decreased amount of T5-1 and 8.1.6 RNA was loaded on the left and centre blots, which highlight reduction in *DMB* RNA in mutants.

METHODS. Mutants 7.2.6, 9.28.3, 9.25.3, 9.10.3, 7.9.6, 7.12.6, 10.66.6, 7.19.6 and 10.78.6 were isolated by the same immunoselection protocol as mutant 9.5.3 (Fig. 1a). Total cellular RNA (generally 20 μ g per lane) was prepared by the guanidinium thiocyanate method²⁶, fractionated by electrophoresis and transferred to Zeta Probe GT (Bio-Rad) membranes. Blots were hybridized according to the manufacturer's guidelines with a full-length *DMB* cDNA (ref. 7) and re-probed with a 540-bp fragment amplified from the β -actin gene (5' primer:



GTGGGGCGCCCCAGGCACCA; 3' primer: CTCCTTAATGTCACGACGATTC; provided by E. Loh). Autoradiography was for 6–7 days with *DMB* probe and 4 h with β -actin probe.

FIG. 3 Mutant 7.12.6 has a milder phenotype, which is associated with a point mutation in the *DMB* gene. **a**, Mutant 7.12.6 binds monoclonal antibody 16.23 at higher levels than class II-presentation-defective mutant 9.10.3. Indirect immunofluorescent analysis was carried out using the antibodies: anti-DR3 (—), 16.23 (ref. 23); anti-DR dimer (· · ·), L243 (ref. 27). Binding of fluoresceinated goat anti-mouse antibody alone was comparable on all cell lines (M. Amaya, not shown). **b**, Proliferation of HLA class II-restricted T cells stimulated by mutant 7.12.6 (■), mutant 9.10.3 (▨) or progenitor 8.1.6 (□) as APC. Polyclonal, class II-restricted T-cell lines are designated by antigen (purified protein derivative of *Mycobacterium tuberculosis* (PPD) or tetanus toxoid (TT)). T-cell lines with the same apparent specificity were derived from separate donors. Specific T-cell stimulation was measured by ^3H -thymidine incorporation, as described². Data are median values from triplicate cultures from a representative experiment; similar results were obtained in three or more independent experiments. Stimulation of T-cells with 8.1.6 as APC: PPD/class II, 6,564 c.p.m.; TT/class II, 14,011 c.p.m.; TT/class II, 8,953 c.p.m.; alloreactive (Allo)/DR3, 17,049 c.p.m.

METHODS. For FACS analysis, cells were incubated with saturating amounts of unlabelled primary antibody, followed by saturating amounts of fluoresceinated goat anti-mouse IgG (H + L) (Gibco-BRL). Stained cells were analysed on an EPICS Elite (Coulter); cell number is displayed against a 4-log-unit axis of fluorescence intensity. The isolation and assay conditions of the antigen-specific T cells have been described²⁸ (the alloreactive anti-DR3 clone²⁹ was a gift from A. Johnson).

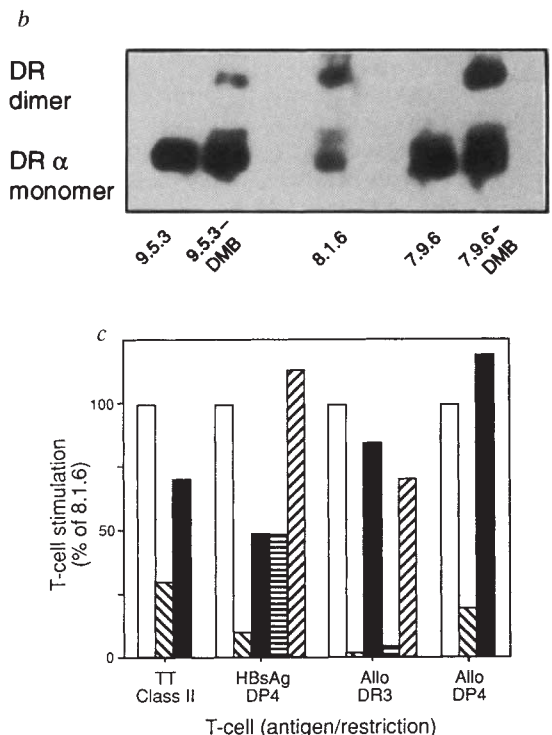
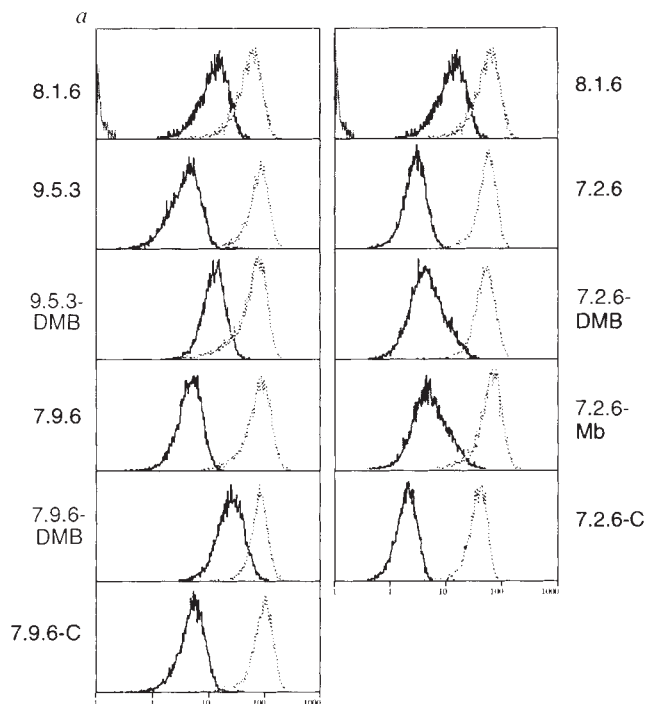
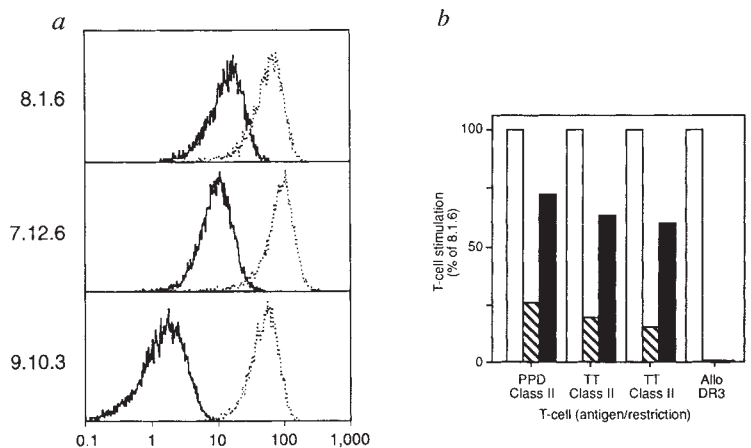


FIG. 4 *DMB* cDNA transfer restores the wild-type phenotype in mutants 9.5.3, 7.9.6 and 7.2.6; cDNA transfer of murine homologue *Mb* is also effective. **a**, Staining of transfectants 9.5.3-DMB, 7.9.6-DMB, 7.2.6-DMB, 7.2.6-Mb and control transfectants 7.9.6-C, 7.2.6-C with the 16.23 antibody. Indirect immunofluorescent analysis was carried out with: Control (---), fluoresceinated goat anti-mouse antibody alone; anti-DR3 (—), 16.23 (ref. 23); anti-DR dimer (· · ·), L243 (ref. 27). **b**, Stability in SDS detergent of HLA-DR dimers from mutants 9.5.3, 7.9.6, progenitor 8.1.6 and transfectants 9.5.3-DMB and 7.9.6-DMB. Unboiled whole-cell extracts were analysed by one-dimensional SDS-PAGE and immunoblotted with monoclonal antibody DA6.147, which recognizes DR dimer and DR α monomers (primary reference cited in ref. 2). **c**, Proliferation of HLA class II-restricted T cells stimulated by mutants 9.5.3 (▨), 7.9.6 (■), their progenitor 8.1.6 (□) or transfectants 9.5.3-DMB (▨) or 7.9.6-DMB (■) as APC. T cells are designated by restricting element and antigen; antigens are tetanus toxoid (TT) and hepatitis B surface antigen (HBsAg). T-cell stimulation was measured as in **b**; stimulation by 8.1.6: TT/class II, 23,108 c.p.m.; HBsAg/DP4, 19,210 c.p.m.; Allo/DR3, 7,030 c.p.m.; Allo/DP4, 50,690 c.p.m. Restoration of stability in SDS and T-cell recognition are incomplete in the transfectants, suggesting that DMB protein levels are below the level of endogenous DMB in the progenitor cell. Nonetheless, staining of the transfectants with the 16.23 antibody is discordantly

high. This discordance is unexplained, but may indicate that only a specific subset of DR-peptide complexes bears the 16.23 epitope in wild-type cells, and that these peptides are more efficiently loaded onto DR molecules under conditions where DM is limiting.

METHODS. Full-length *DMB* and *Mb* cDNA clones^{6,7} were inserted into the expression vector pcDNA1/NEO (Invitrogen). Mutant lymphoblastoid cell lines were electroporated (Gene-pulser, Bio-Rad) with *Scal*-linearized plasmid DNA. After 3–4 days, selection was initiated in $1\text{--}1.5\text{ mg ml}^{-1}$ G418 (Gibco-BRL). Indirect immunofluorescence was carried out as in **a**; staining with control antibody was approximately equivalent for all cells and is shown for progenitor 8.1.6 only. Western blot analyses were carried out essentially as described², except that polyvinylidene difluoride membranes were used and bands were developed with enhanced chemiluminescent reagents (Gibco-BRL). We often observe a higher proportion of stable class II dimers in progenitor 8.1.6 than is shown in **b**; some variability apparently results from varied growth conditions of the cells (data not shown). The isolation of the DP4-restricted, HBsAg-specific T cells has been described². The alloreactive, anti-DP4 T cells³⁰ were provided by S. Masewicz. Recombinant HBsAg was a gift from Merck, Sharp and Dohme and was used at $1\text{ }\mu\text{g ml}^{-1}$. T-cell proliferation was assayed as in **b**.

domain also shows limited polymorphism (in Ma/Mb; E. Hermel and J.M., unpublished data) and contains seven cysteine residues, five in novel positions^{6,7}.

The failure of the *DMB* mutants to generate the normal repertoire of MHC class II-peptide complexes suggests several models for DM function. DM could act as a class II chaperone, mediating either a trafficking step or a folding event necessary for intracellular peptide binding. If HLA-DM assumes a class II-like structure, however, DM may instead act as a shuttle to deliver peptides to a compartment for association with conventional class II molecules. In this scenario, peptide binding to DM constitutes an initial step in determinant selection for class II-binding peptides, and the association of class II molecules with Ii fragments in the mutants reflects decreased availability of other peptides. Alternatively, the DM molecule may be required for optimal removal of the Ii peptides (residues 80–103). In this model, the peptide-binding groove of DM has evolved to bind Ii peptides with high efficiency. The existence of DM as a 'sink' for the Ii peptide would remove Ii without the use of peptidases in a compartment probably also containing immunogenic peptides.

In these models, we postulate that HLA-DM is a peptide-binding molecule with a fundamentally different role from that of other class II molecules: the peptide-DM complex is not a ligand for T cells and its function is intracellular. Consistent with the possibility that DM does not interact with T cells, the DMB (and Mb) sequence lacks homology to other class II β -chains in the conserved region of the β 2 domain which mediates interaction with the CD4 co-receptor^{16,17}. Also, mice that lack I-A and I-E molecules, but have normal *Ma* and *Mb* genes, have almost no mature CD4⁺ T cells, suggesting that *Ma*/*Mb* is not a positive selecting element for thymic T cells^{18,19}. Consistent with the possibility that DM functions intracellularly, the DMB and Mb sequences also diverge from class II β -chains at three of four highly conserved positions (80–83) that are implicated in transport and targeting of class II molecules to the cell surface^{20,21}. In the DMB and Mb, threonine substitutes for asparagine in this region (position 82N in I-A β); a similar substitution (serine for asparagine) in mutant I-A β -chains results in intracellular retention of sialylated I-A molecules²⁰.

In related work, we isolated a *DRA* point mutant in which aberrant glycosylation in the second domain of HLA-DR α results in association of the mutant DR molecules with Ii peptides (amino acids 80–103)⁹. As mutation of both *DRA* and *DMB* generates HLA-DR molecules with similar characteristics, one attractive hypothesis is that direct association between class II molecules and DM molecules is required for DM function. In this regard, homotypic interaction between HLA-DR1 molecules occurs during crystallization and involves an interface near the region altered in the mutant DR molecules¹⁵.

Our findings indicate that MHC class II-peptide association in B-lymphoblastoid cells is significantly modulated by the level of expression of HLA-DM. The co-regulated expression of DM and other class II molecules (P.M., unpublished results, and ref. 7), and their coordinated induction by γ -interferon⁷, probably promote presentation of non-self peptides by activated APC. Conversely, lack of DM expression favours presentation of a limited range of peptides; thus, it will be of interest to determine whether conventional class II molecules are expressed in the absence of DM in any normal cells, where this discordance of expression might have an immunoregulatory role. □

8. Pious, D., Dixon, L., Levine, F., Cotner, T. & Johnson, R. *J. exp. Med.* **162**, 1193–1207 (1985).
9. Mellins, E. et al. *J. exp. Med.* **179**, 541–549 (1994).
10. Riberdy, J. M., Newcomb, J. R., Surman, M. J., Barbosa, J. A. & Cresswell, P. *Nature* **360**, 474–477 (1992).
11. Sette, A. et al. *Science* **258**, 1801–1804 (1992).
12. Hunt, D. F. et al. *Science* **258**, 1817–1820 (1992).
13. Chiciz, R. M. et al. *J. exp. Med.* **178**, 27–47 (1993).
14. Spies, T. et al. *Nature* **348**, 744–747 (1990).
15. Brown, J. H. et al. *Nature* **364**, 33–39 (1993).
16. Cammarota, G. et al. *Nature* **356**, 799–801 (1992).
17. Koniig, R., Huang, L.-Y. & Germain, R. N. *Nature* **356**, 796–798 (1992).
18. Cosgrove, D. et al. *Cell* **66**, 1051–1066 (1991).
19. Grusby, M. J., Johnson, R. S., Papaioannou, V. E. & Glimcher, L. H. *Science* **253**, 1417–1420 (1991).
20. Griffith, I. J. et al. *J. exp. Med.* **167**, 541–555 (1988).
21. Sant, A. J. & Chervonsky, A. *Hum. Immun.* **37**, 3 (1993).
22. Campbell, R. D. & Trowsdale, J. *Immun. Today* **14**, 349–352 (1993).
23. Johnson, J. P., Meo, T., Riethmuller, G., Schendel, D. J. & Wank, R. *J. exp. Med.* **156**, 104–111 (1982).
24. Gladstone, P., Furesz, L. & Pious, D. *Proc. natn. Acad. U.S.A.* **79**, 1235–1239 (1982).
25. Gladstone, P. & Pious, D. *Nature* **271**, 459–461 (1978).
26. Xie, W. & Rothblum, L. I. *BioTechniques* **11**, 325–327 (1991).
27. Lampson, L. A. & Levy, R. J. *Immun.* **125**, 293–299 (1980).
28. Mellins, E., Woelfel, M. & Pious, D. *Hum. Immun.* **18**, 211–223 (1987).
29. Johnson, A. H. et al. in *Immunobiology of HLA* (ed. Dupont, B.O.) 202–203 (Springer, New York, 1984).
30. Shaw, S., Johnson, A. & Shearer, G. *J. exp. Med.* **152**, 565–580 (1980).

ACKNOWLEDGEMENTS. P.M. and J.S. contributed equally to this study. The isolation and initial phenotypic characterization of these mutants was carried out by E.M. and others in the laboratory of D. Pious. We thank D. Chaplin, M.-Y. Chang and E. Loh for PCR primers, and D. Wiley and J. Brown for discussion. This work was supported by grants from the NIH (E.M., J.M.) and an Immunology Training grant to the University of Pennsylvania (J.M. and Monsanto (E.M.).

HLA-DMA and -DMB genes are both required for MHC class II/peptide complex formation in antigen-presenting cells

Steven P. Fling*, Benjamin Arp* & Donald Plous*^{†‡}

Departments of * Pediatrics, † Immunology and ‡ Genetics, University of Washington, Seattle, Washington 98195, USA

MAJOR histocompatibility complex (MHC) class II molecules are highly polymorphic cell-surface glycoproteins that present antigenic peptides to CD4⁺ T lymphocytes. The normal assembly of class II molecules with cognate peptides for antigen presentation requires an accessory function provided by a gene mapping to the class II region of the HLA complex^{1,2}. The isolation of somatic cell mutants of antigen-presenting cells (APC) has shown that at least one gene which maps between HLA-DP and HLA-DQ, provisionally designated c2p-1 (ref. 3), mediates this process^{1,2,4}. Here we describe a unique new mutant 2.2.93, which manifests defective formation of class II/peptide complexes like that described in c2p-1 mutants. We show that (1) mutant 2.2.93 contains a mutation in HLA-DMA⁵, and a representative c2p-1 mutant, 9.5.3, contains a mutation in HLA-DMB⁵; and (2) transfection and expression of DMA complementary DNA in 2.2.93, and DMB cDNA in 9.5.3, reverses their mutant phenotypes. These results show that HLA-DMA and -DMB, genes of previously unknown function mapping between HLA-DP and HLA-DQ⁵, are required for the normal assembly of peptides with MHC class II molecules. They suggest that HLA-DMA and -DMB encode subunits of a functional heterodimer which is critical in the pathway of class II antigen presentation.

Human B-lymphoblastoid cell lines (B-LCL) containing deletions within the HLA region have been used to derive mutants that define a gene required for class II restricted antigen presentation⁶. These mutants fail to assemble normal MHC class

Received 18 October 1993; accepted 12 January 1994.

1. Rudensky, A. Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. & Janeway, C. A. Jr *Nature* **353**, 622–626 (1991).
2. Mellins, E. et al. *Nature* **343**, 71–74 (1990).
3. Mellins, E., Kempin, S., Smith, L., Monji, T. & Pious, D. *J. exp. Med.* **174**, 1607–1615 (1991).
4. Riberdy, J. M. & Cresswell, P. *J. Immun.* **148**, 2586–2590 (1992).
5. Ceman, S., Rudersdorf, R., Long, E. O. & DeMars, R. *J. Immun.* **149**, 754–761 (1992).
6. Cho, S., Attaya, M. & Monaco, J. J. *Nature* **353**, 573–576 (1991).
7. Kelly, A. P., Monaco, J. J., Cho, S. & Trowsdale, J. *Nature* **353**, 571–573 (1991).

‡ To whom correspondence should be addressed.

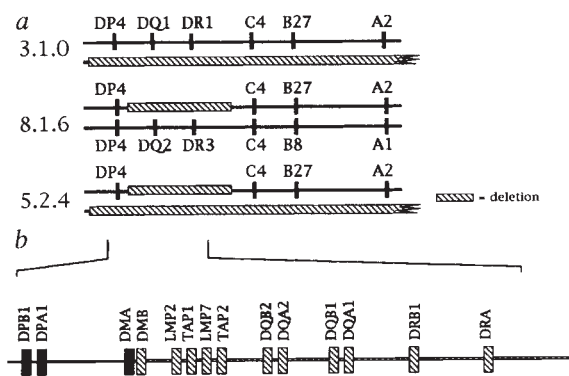
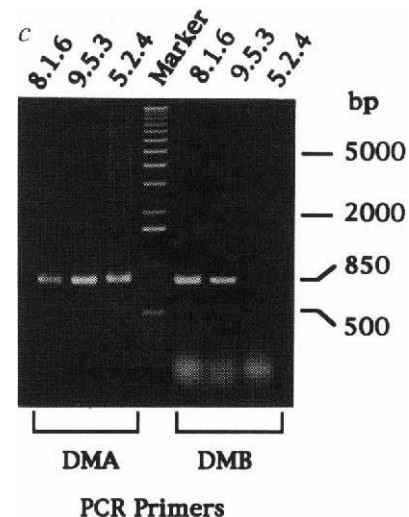


FIG. 1 a, The HLA portions of the hemizygous deletion in 3.1.0, the progenitor of 2.2.93; of the HLA hemizygous deletion in 8.1.6, the progenitor of 9.5.3; and of the homozygous deletion in 5.2.4. Deletions are indicated by horizontal hatched bars on the deleted haplotype and have been described for 3.1.0 (refs 8, 16), 8.1.6 (ref. 8), and 5.2.4 (refs 1, 7). The centromere is to the left. b, Details of the HLA class II region (based on ref. 17). Hatched boxes indicate loci homozygously deleted in 5.2.4. c, PCR analysis of DMA and DMB gene expression in 5.2.4. cDNAs from the indicated cell lines were used in separate reactions with DMA and DMB primers. The 859 bp DMA band is present in all three cell lines. The 847 bp DMB band is present for 8.1.6 and 9.5.3, but not for 5.2.4. Selected sizes of the 1 kb ladder DNA Markers (Gibco BRL) are indicated in basepairs (bp).

METHODS. Mutant 2.2.93 was isolated from ethyl methane sulphonate mutagenized 3.1.0/DR3 by immunoselection⁹ with anti-DR3 antibody 7.3.19.1 (ref. 18) and rabbit complement. To derive 3.1.0/DR3, a 1.2 kb DR3B1 cDNA was cloned from B-LCL 8.1.6 and subcloned into pLNCx, a retroviral expression vector¹⁹. This was used to obtain clones of the packaging line PA317 (ref. 19) and to transduce B-LCL 3.1.0. G418



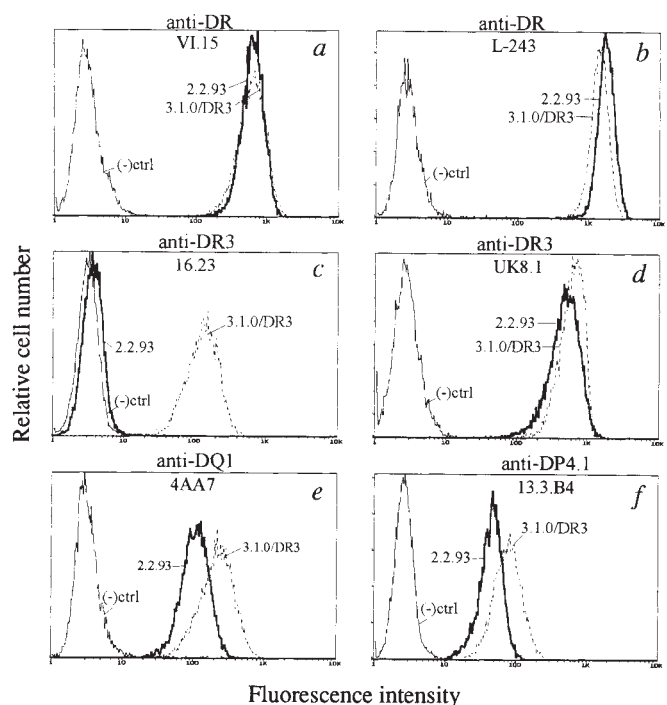
resistant cultures of transduced 3.1.0 were cloned and screened for stable DR3 expression to obtain 3.1.0/DR3. For PCR analysis, RNAs were purified from mutant and progenitor cells and cDNAs were prepared using M-MLV reverse transcriptase (BRL) under standard first-strand reaction conditions. In PCR reactions, the following primers were used: for DMB, primers DMB-5 (plus strand, 5' UT region, 5'-ATCTTT-ACAGAGCAGAGCATGA-3') and DMB-3M (minus strand, 3' UT region, 5'-CCTTCTCACTGGAGTGAAGTTG-3') were used to generate a 847 bp DNA fragment; for DMA, primers DMA-55 (plus strand, 5' UT region, 5'-CTAAAGCTGGGTTGGTAGCT-3') or DMA-5 (plus strand, 5' UT region, 5'-CCTACTGTGTGGCAAGAAGGT-3'), and DMA-3M (minus strand, 3' UT region, 5'-GCTGGCATCAAACTCTGGTCT-3') were used to generate an 859 bp and an 835 bp DNA fragment, respectively. Standard PCR conditions were used. Samples were run on 1% agarose/TAE minigels.

FIG. 2 Mutant 2.2.93 expresses conformationally altered class II molecules. Mutant 2.2.93 (dark lines), its progenitor 3.1.0/DR3 (dashed lines), and negative controls (thin lines) were stained by indirect immunofluorescence using antibodies: a, VI.15 (ref. 16) and b, L243 (ref. 20) to HLA-DR; c, 16.23 (ref. 21) and d, UK8.1 (ref. 22) to HLA-DR3; e, 4AA7 (ref. 23) to HLA-DQ1; and f, 13.3.B4 (ref. 24) to HLA-DP4.1. Negative controls: a, 2.11.93 (3.1.0/DR3-derived DR-null mutant; unpublished results); b, e, 2.4.93 (3.1.0/DR3-derived DR/DQ null mutant; ref. 3 and unpublished results); c, d, 3.1.0 (refs 8, 16); f, 6.1.6 (class II null mutant²⁵).

METHODS. For immunofluorescent flow cytometry, 0.5×10^6 cells were stained for 1 h at 4 °C with saturating amounts of the indicated antibodies in 50 μ l RPMI, 5% FBS, washed and further stained for 1 h at 4 °C with 1:100 fluoresceinated goat F(ab')₂ anti-mouse IgG Fc (FITC-GAM) (Sigma). After staining, cells were washed, resuspended in 25 μ g ml⁻¹ propidium iodide (PI) for dye exclusion (viability) gating, and run through FACScan (Becton-Dickinson) and analysed with REPRO-MAN. 3.1.0/DR3 and 2.2.93 cells in each panel were analysed in parallel. Histograms in Figs 2, 3 and 4 represent immunofluorescence on viable, triple-gated events (forward scatter FSC, side scatter SSC, and PI exclusion).

II/peptide complexes, resulting in conformational changes in class II molecules and defective antigen presentation^{1,6}. Complementation analysis suggests that a single gene, c2p-1, mapping to a 230 kilobase (kb) region between HLA-DP and HLA-DQ delimited by homozygous deletions in B-LCL 5.2.4 (ref. 7) and B-LCL 721.82 is responsible for this phenotype^{1,3}.

To derive mutant 2.2.93, we used as a progenitor B-LCL 3.1.0, which contains a ~40 megabase (Mb) hemizygous deletion in chromosome 6p, removing one HLA haplotype⁸ (Fig. 1a). To exploit conformational changes in DR3 class II molecules as a basis for selection of this type of mutant⁶, 3.1.0 (DR1 + /DR3 -) was transfected with DR3B1 (Fig. 1 legend). The resulting 3.1.0/DR3 progenitor was mutagenized and immunoselected⁹ with monoclonal antibody 7.3.19.1, which exhibits reduced binding



on c2p-1 mutants as a result of conformational changes in HLA-DR3 molecules⁶. Surviving clones were analysed for class II conformational alterations indicative of defects in class II/peptide assembly.

To assess the conformational state of cell surface class II molecules in mutant 2.2.93, we measured binding of antibodies to DR, DP and DQ determinants. The defect in class II/peptide assembly in c2p-1 mutants, exemplified by mutant 9.5.3 (refs

6, 10), is characterized in part by normal levels of surface class II molecules with altered conformation. This is manifested by diminished binding of certain antibodies specific for polymorphic class II determinants⁶. Mutant 2.2.93, like 9.5.3, expresses levels of class II molecules comparable to its progenitor (Fig. 2a, b), but has severely diminished expression of the DR3 determinant recognized by antibody 16.23 (Fig. 2c). The binding of the anti-DR3,5,6 antibody UK8.1 (Fig. 2d) verifies that loss of the 16.23 determinant is not caused by loss of DR3B1 transgene expression. Furthermore, although 2.2.93 was immunoselected with an antibody against DR3, the class II conformational changes affect not only determinants on DR3, but also those recognized by DQ1-specific antibody 4AA7 (Fig. 2e) and DP4.1-specific antibody 13.3.B4 (Fig. 2f).

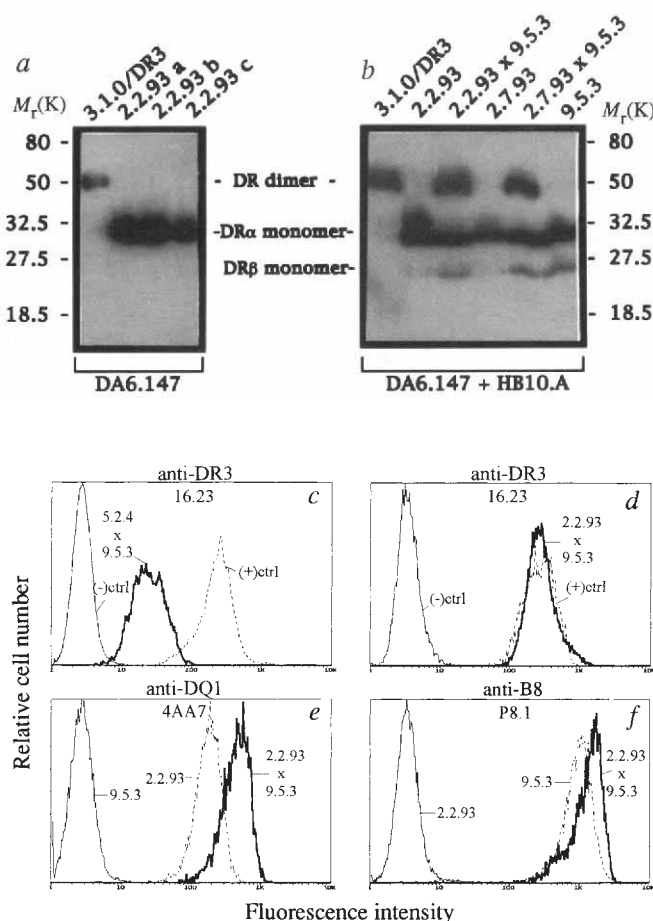
We compared the stability in sodium dodecyl sulphate (SDS) of class II dimers extracted from mutant 2.2.93 and progenitor 3.1.0/DR3. The stability of class II dimers in SDS reflects the state of peptide binding in MHC class II/peptide complexes^{4,11,12}. The defective assembly of class II/peptide complexes in c2p-1 and similar mutants results in the instability of dimers when run in SDS-polyacrylamide gel electrophoresis (SDS-PAGE)^{1,4,6}. Like class II DR dimers from c2p-1 mutants, dimers from 2.2.93 dissociate completely into DR α and DR β monomers in SDS-PAGE, whereas dimers extracted from 3.1.0/DR3 remain stable (Fig. 3a). Similarly, DP and DQ dimers from 2.2.93 dissociate in SDS-PAGE, whereas those from 3.1.0/DR3 remain intact (not shown). These changes in dimer stability and the characteristic changes in binding of certain antibodies to DR, DQ and DP demonstrate that the defect in mutant 2.2.93 results in the expression of conformationally altered class II molecules and affects all conventional class II molecules. This

is similar to the defective assembly of MHC class II/peptide complexes in c2p-1 mutants.

To determine if the phenotypic similarities between 2.2.93 and c2p-1 mutants are caused by mutations in the same or different genes, we made somatic cell hybrids between 2.2.93 and 9.5.3. The mutations in 2.2.93 and 9.5.3 are complementary, as shown by restoration to wild-type levels of antibody 16.23 staining on 2.2.93 $\times$ 9.5.3 hybrids (Fig. 3c-f) and by the appearance of SDS-stable class II dimers in cell lysates from these hybrids (Fig. 3b). These results contrast with results from hybrids between c2p-1 mutants, in which these defects are non-complementary (ref. 1; Fig. 3c). Because all c2p-1 mutants, including 5.2.4, are derived from the same progenitor, 8.1.6 (ref. 8; Fig. 1a), the fact that they are non-complementary suggested that a single gene within the region which is homozygously deleted in 5.2.4 (ref. 7; Fig. 1a) could effect this mutant phenotype¹. We postulated that a mutation in a second gene outside the 5.2.4 deletion must be responsible for the defect in 2.2.93 and, because of the phenotypic similarities between 2.2.93 and c2p-1 mutants, that the two genes encode products which interact.

To investigate this possibility, we analysed the c2p-1 mutant 5.2.4 and its progenitor 8.1.6 for the expression of DMA and DMB, two genes mapping near the centromeric breakpoint of the 5.2.4 homozygous deletion. This breakpoint had previously been mapped between Y4 (RING3) and Y3 (TAPI)⁷. As assessed by polymerase chain reaction (PCR), 8.1.6 expresses both DMA and DMB, whereas 5.2.4 expresses DMA but does not express DMB (Fig. 1c). These results suggested that the hemizygous deletion in 8.1.6 and the homozygous deletion in 5.2.4 extend into the DMB gene but not into the more centromeric DMA gene (Fig. 1b). We therefore reasoned that HLA-

FIG. 3 Stability of HLA-DR dimers from mutant 2.2.93 and 2.2.93 $\times$ 9.5.3 hybrids; and cell-surface expression of MHC molecules on 2.2.93 $\times$ 9.5.3 hybrids. a, Mutant 2.2.93 expresses unstable class II dimers. Cell lysates from 3.1.0/DR3 and three independent lysates of mutant 2.2.93 were run in dimer stability assays and were stained with antibody DA6.147, which recognizes HLA-DR α monomers and HLA-DR dimers²⁶. b, The defects in dimer stability in mutants 2.2.93 and 9.5.3 are complementary. Cell lysates from the indicated cells and hybrids were run in dimer stability assays and were stained with a combination of antibody DA6.147 and HB10.A. HB10.A reacts with HLA-DR3 β in monomers and in HLA-DR α / β dimers²⁷. Another mutant, 2.7.93, is included but has not been characterized. In progenitor cells, the percentage of DR subunits that form dimers ranges from 50 to 100% (refs 6, 10). c-f, Immunofluorescence of 2.2.93 $\times$ 9.5.3 hybrids. c, c2p-1 mutants are non-complementary¹. Negative control 5.2.4 (thin line), negative control hybrid 5.2.4 $\times$ 5.2.4 (dashed line), and positive control hybrid 8.1.6 $\times$ 5.2.4 (dashed line) were stained using antibody 16.23 followed by FITC-GAM. d, Mutants 2.2.93 and 9.5.3 are complementary. Negative control 2.2.93 (thin line), hybrid 2.2.93 $\times$ 9.5.3 (dark line), and positive control hybrid 2.4.93 $\times$ 9.5.3 (dashed line) were stained using antibody 16.23 followed by FITC-GAM. Positive control hybrid 2.4.93 $\times$ 9.5.3 was made with 2.4.93 (3.1.0/DR3-derived DR/DQ null mutant; ref. 3 and unpublished results). It was used to determine the levels of antibody 16.23 immunofluorescence expected in hybrids made with a 3.1.0/DR3-derived positive control. e, f, Verification of the HLA haplotypes in 2.2.93 $\times$ 9.5.3 hybrids. 2.2.93 $\times$ 9.5.3 hybrids (dark lines) and controls 2.2.93 and 9.5.3 were stained using antibodies: e, 4AA7 (HLA-DQ1); and f, P8.1 (HLA-B8, ref. 28) followed by FITC-GAM. Staining with antibodies 4AA7 and P8.1 confirms retention of chromosome 6p haplotypes from both partners (see Fig. 1a) in 2.2.93 $\times$ 9.5.3 hybrids. The tetraploid status of the hybrids was verified by DNA content determination using PI (not shown). METHODS. For dimer stability assays, unboiled, detergent (NP-40)-extracted cell lysates were prepared and 1×10^6 cell equivalents per sample were run in non-reducing SDS-PAGE and transferred as described¹⁰. Membranes were stained with the indicated antibodies followed by polyclonal rabbit anti-mouse IgG (Sigma) and ¹²⁵I-labelled protein-A. Somatic cell hybrids were prepared by polyethylene glycol (PEG) fusion as described⁶ and were selected using methotrexate (2×10^{-7} M) and G418 (1 mg ml^{-1}). Hybrids are stable lines. For flow cytometry, cells were stained and analysed as described in Fig. 2.



DMB was a candidate gene for the c2p-1 locus, and that a defect in HLA-DMA in mutant 2.2.93 could account for both its phenotype and the 2.2.93 $\times$ 9.5.3 complementation.

HLA-DMA cDNAs from 2.2.93 and 3.1.0, and DMB cDNA from 9.5.3 were PCR-amplified and sequenced. Although the sequence of DMA from 3.1.0 is normal⁵, in the sequence of DMA from 2.2.93 there is a single nucleotide substitution which changes the codon at amino-acid position 81 from TGG to TGA. This change substitutes a stop codon for the tryptophan codon in the amino-terminal portion of HLA-DMA. In the sequence of DMB from 9.5.3, there is a single nucleotide substitution which changes the codon at amino-acid position 120 from TGG to TGA. This change introduces a stop codon in place of the tryptophan codon in the predicted β 2-like domain of DMB⁵.

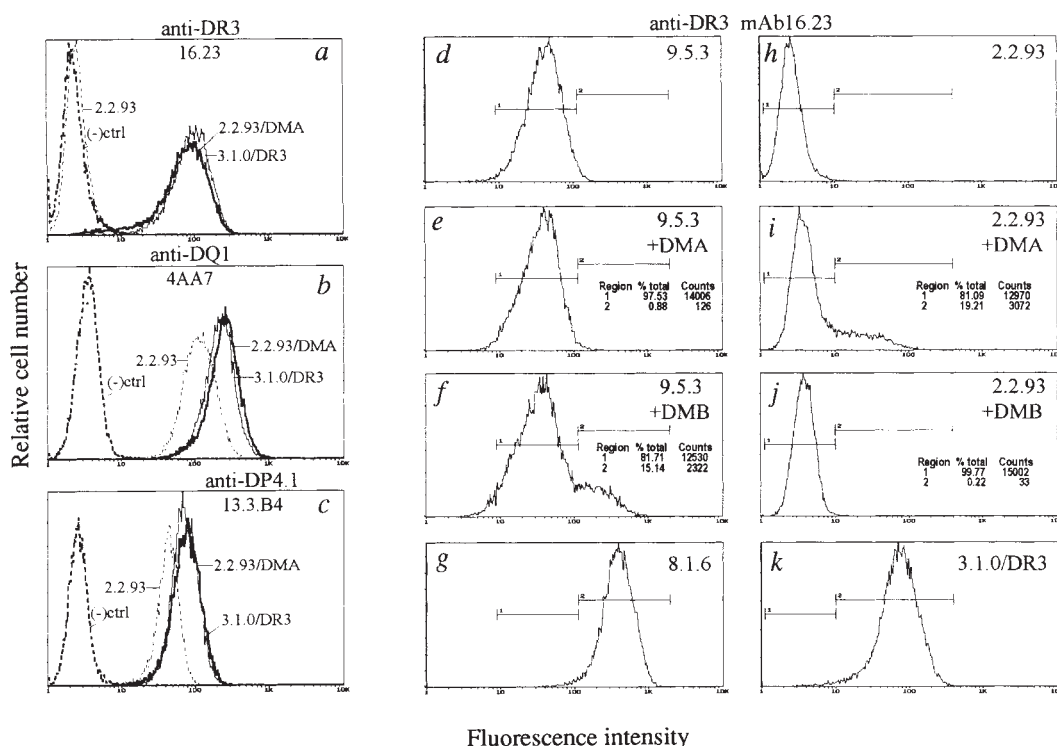
We next determined if normal HLA-DMA cDNA derived from 3.1.0 could complement the defect in 2.2.93. Stable transduction of DMA into 2.2.93 restores to wild-type levels the conformation-sensitive DR, DQ and DP determinants recognized by antibodies 16.23, 4AA7, and 13.3.B4, respectively (Fig. 4a-c) and results in the appearance, in cell lysates, of stable class II dimers in SDS-PAGE (not shown). We then investigated whether DMB cDNA derived from c2p-1 progenitor 8.1.6 could complement the defect in 9.5.3 as assessed by restoration of antibody 16.23 staining in 9.5.3 cells transiently transfected with DMB. Electroporation of 9.5.3 with DMB cDNA induced antibody 16.23 staining in up to 14% of the cells (Fig. 4f). As

controls, transfections of DMA cDNA did not induce 16.23 staining in 9.5.3 (Fig. 4e). Additionally, DMA cDNA, but not DMB cDNA, induced antibody 16.23 staining in 2.2.93 (Fig. 4i, j) in up to 19% of the cells in transient transfections. These results are consistent with the sequence data and the 2.2.93 $\times$ 9.5.3 complementation.

We have described a unique mutant, 2.2.93, with a defect in the assembly of MHC class II/peptide complexes, and have shown that the defective gene is HLA-DMA. We have also shown that HLA-DMB is the defective gene in 9.5.3, a previously described c2p-1 mutant with a similar defect which results in the inability to present native antigens⁶. HLA-DMA and HLA-DMB are genes of previously unknown function mapping to the class II region of the MHC⁵. Our findings show that HLA-DM provides a critical accessory function in the generation of normal MHC class II/peptide complexes. The precise role of HLA-DM in this process is unknown.

The fact that genetically distinct lesions in DMA and DMB result in a similar phenotype suggests that DMA functions in association with DMB. The generalized amino-acid homology of DMA with class I and class II A genes and of DMB with class I and class II B genes is consistent with the hypothesis that DMA and DMB, like other class I and class II gene products, pair as heterodimers. Classical class II DR, DP and DQ molecules associate with invariant chain (Ii) and are transported to endocytic compartments. After Ii dissociation, class II hetero-

FIG. 4 Cell-surface expression of conformationally normal class II molecules is restored in 2.2.93 by transfection of DMA, and in 9.5.3 by transfection of DMB. a-c, Mutant 2.2.93 (thin-dashed lines), progenitor 3.1.0/DR3 (thin lines), negative controls (heavy-dashed lines), and 2.2.93/DMA (2.2.93 stably transduced with DMA) (dark lines) were stained using antibodies: a, 16.23 (DR3); b, 4AA7 (DQ1); and c, 13.3.B4 (DP4.1) followed by FITC-GAM. Negative controls: a, c, 6.1.6 (class II null mutant²⁵); and b, 9.5.3. d-k, Transient transfections of DMA and DMB. Mutant 9.5.3 (d-f) and its progenitor 8.1.6 (g) and mutant 2.2.93 (h-j) and its progenitor 3.1.0/DR3 (k) were stained with antibody 16.23 followed by FITC-GAM. Mutants were untransfected (d, h) or were transfected with plasmid DMA (pDMA) (e, i) or pDMB (f, j). In 12 experiments, the median per cent of pDMB-transfected 9.5.3 cells staining positive for the 16.23 epitope above control (e) is 4.7 ± 1.1 s.e.m. Similar results are obtained with 2.2.93 transfected with pDMA (i; n=8). The % positive cells are consistent with frequencies of positive cells in other transient expression systems²⁹. Complementation is specific for the 16.23 determinant: 9.5.3 cells transfected with pDMB and stained with antibody 4AA7 were negative, as were 2.2.93 cells transfected with pDMA and stained with antibody P8.1 (data not shown). METHODS. For stable transfectants, an 835 bp cDNA encoding HLA-DMA was cloned from B-LCL 3.1.0 (Fig. 1) and subcloned into pLXSHD, a retroviral expression vector³⁰ containing *hisD*. This construct was used to derive clones of the packaging line PA317 (ref. 19) and



to transduce 2.2.93. The 2.2.93/DMA lines were obtained by selection of transduced 2.2.93 in 4 mM L-histidinol (Sigma). a-c are from one line and are representative of four independent lines. For construction of transient expression plasmids, an 847 bp cDNA encoding HLA-DMB from 8.1.6, and an 835 bp cDNA encoding HLA-DMA from 3.1.0 were cloned (Fig. 1) into the expression vector pcDNA1/AMP (Invitrogen) to obtain plasmids pDMB and pDMA, respectively. pcDNA1/AMP contains Polyoma and SV40 origins of replication allowing efficient episomal replication. For transient transfections, 2.5×10^7 cells were resuspended in 0.5 ml RPMI with 10–100 μ g circular pDMB or pDMA and were electroporated using 2.2 kV. Samples were cultured for 72 h to regain viability and then stained and analysed by flow cytometry as described in Fig. 2.

dimers are normally free to bind peptides from processed antigens. Interestingly, c2p-1 mutants (ref. 13; T. Monji and D.P., manuscript in preparation) and phenotypically similar mutants^{14,15} exhibit an aberrant association of HLA-DR molecules with Ii-derived peptides. Whether the mutation in 2.2.93 results in the same aberrant association of Ii peptides with DR molecules is under investigation. Given the similarities in primary structure of DMA and DMB to classical MHC mol-

ecules which function to bind peptides for presentation, a DM heterodimer could bind peptides. Modelling suggests DM may have an antigen-binding cleft similar to HLA-DR but with a restricted spectrum of peptide binding⁵. An HLA-DM heterodimer may therefore function as a receptor for these Ii peptides, freeing DR, DQ and DP molecules to bind cognate peptides for antigen presentation. □

Received 20 December 1993; accepted 7 March 1994.

1. Mellins, E., Kempin, S., Smith, L., Monji, T. & Pious, D. *J. exp. Med.* **174**, 1607–1615 (1991).
2. Ceman, S., Rudersdorf, R., Long, E. O. & DeMars, R. *J. Immun.* **149**, 754–761 (1992).
3. Pious, D., Fling, S. & Monji, T. in *Antigen Processing and Presentation* (eds Humphries, R. & Pierce, S.) (Academic, New York, in the press).
4. Riberdy, J. M. & Cresswell, P. *J. Immun.* **148**, 2586–2590 (1992).
5. Kelly, A. P., Monaco, J. J., Cho, S. & Trowsdale, J. *Nature* **353**, 571–573 (1991).
6. Mellins, E. *et al. Nature* **343**, 71–74 (1990).
7. Spies, T. *et al. Nature* **348**, 744–747 (1990).
8. Levine, F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3741–3745 (1985).
9. Pious, D., Soderland, C. & Gladstone, P. *Immunogenetics* **4**, 437–448 (1977).
10. Pious, D., Dixon, L., Levine, F., Cotner, T. & Johnson, R. *J. exp. Med.* **162**, 1193–1207 (1985).
11. Germain, R. N. & Hendrix, L. R. *Nature* **353**, 134–139 (1991).
12. Sadegh-Nasserli, S. & Germain, R. N. *Nature* **353**, 167–170 (1991).
13. Mellins, E. *J. exp. Med.* **179**, 541–549 (1993).
14. Sette, A. *et al. Science* **258**, 1801–1804 (1992).
15. Riberdy, J. M., Newcombe, J. R., Surman, M. J., Barbosa, J. A. & Cresswell, P. *Nature* **360**, 474–477 (1992).
16. Gladstone, P., Furesz, L. & Pious, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1235–1239 (1982).
17. Campbell, D. R. & Trowsdale, J. *Immun. Today* **14**, 349–352 (1993).
18. Koning, F., Schrueder, I., Giphard, M. & Bruning, H. *Hum. Immun.* **9**, 221–230 (1984).
19. Miller, A. D. & Rosman, G. J. *BioTechniques* **7**, 980–990 (1989).
20. Lampson, L. A. & Levy, R. J. *Immun.* **125**, 293–299 (1980).
21. Johnson, J. *et al. J. exp. Med.* **156**, 104–111 (1982).
22. Bodmer, J. G., Heyes, J. M. & Lindsay, J. in *Histocompatibility Testing* (eds Albert, E. D., Baur, M. P. & Mayr, W. R.) (Springer, Berlin, 1985).
23. Clark, E. A. *et al. J. Immun.* **143**, 3873–3880 (1989).
24. Viken, H. D., Gaudernack, G. & Thorsby, E. *Tiss. Antigens* **34**, 250–259 (1989).
25. Levine, F., Erlich, H. A., Mach, B. & Pious, D. *J. Immun.* **134**, 637–640 (1985).
26. Guy, K. *et al. Eur. J. Immun.* **12**, 942–948 (1982).
27. Clark, E. A. & Yakoshi, T. in *Leucocyte Typing* (eds Bernard, A., Baunsell, L., Dausset, J. & Schlossman, S.) (Springer, Heidelberg, 1984).
28. Antonelli, P., Nisperos, B., Braun, M. & Hansen, J. A. *Hum. Immun.* **11**, 11–18 (1984).
29. Margolskee, R., Kavathas, P. & Berg, P. *Molec. cell. Biol.* **8**, 2837–2842 (1988).
30. Stockschiader, M. A. R., Storb, R., Osborne, W. R. A. & Miller, A. D. *Hum. Gene Ther.* **2**, 33–39 (1991).

ACKNOWLEDGEMENTS. S.P.F. and B.A. contributed equally to this work. We thank T. Cotner, J. Rak, T. Monji, S. Fujihara, D. Hill, L. Smith and K. Ryan for contributions and support, J. Trowsdale for communicating data before publication and W. Kwok and W. Osborne for pLNCx, PA317 and pLXSHD. This work was supported by research grants from the NIH, and by a Cancer Research Institute postdoctoral fellowship (S.P.F.).

B61 is a ligand for the ECK receptor protein-tyrosine kinase

Timothy D. Bartley*, Robert W. Hunt*, Andrew A. Welcher*, William J. Boyle*, Vann P. Parker*, Richard A. Lindberg*, Hsieng S. Lu*, Anne M. Colombero*, Robin L. Elliott*, Brenda A. Guthrie*, Paige L. Holst*, James D. Skrine*, Robert J. Toso*, Ming Zhang*, Elsa Fernandez*, Geraldine Trall*, Brian Varnum†, Yosef Yarden†, Tony Hunter‡ & Gary M. Fox*

* Amgen Inc., Amgen Center, Thousand Oaks, California 91320-1789, USA

† Department of Chemical Immunology, Weizmann Institute of Science, Rehovot 76100, Israel

‡ Molecular Biology and Virology Laboratory, The Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92186-5800, USA

A PROTEIN ligand for the ECK¹ receptor protein-tyrosine kinase has been isolated by using the extracellular domain (ECK-X) of the receptor as an affinity reagent. Initially, concentrated cell culture supernatants were screened for receptor binding activity using immobilized ECK-X in a surface plasmon resonance detection system^{2,3}. Subsequently, supernatants from selected cell lines were fractionated directly by receptor affinity chromatography, resulting in the single-step purification of B61, a protein previously identified as the product of an early response gene induced by tumour necrosis factor- α ⁴. We report here that recombinant B61 induces autophosphorylation of ECK in intact cells, consistent with B61 being an authentic ligand for ECK. ECK is a member of a large orphan receptor protein-tyrosine kinase family headed by EPH⁵, and we suggest that ligands for other members of this family will be related to B61, and can be isolated in the same way.

The ECK protein-tyrosine kinase is an orphan receptor originally identified by complementary DNA cloning from a human epithelial cell library¹. Although ECK expression is highest in cells of epithelial origin, the physiological significance of this receptor is unknown. Purified recombinant extracellular domain of ECK (ECK-X), truncated at Asn 534 and produced⁶ in Chinese hamster ovary (CHO) cells, was covalently coupled to the dextran surface of a BIAcore^{2,3} sensorchip and used to screen concentrated cell culture supernatants for receptor binding activity. Several supernatants contained material which bound to the ECK-X surface (Fig. 1). Conditioned medium from SK-BR-3 or HCT-8 cells was concentrated 10–40 times, diafiltered, and loaded directly onto columns of immobilized ECK-X. The pH 4.0 eluate of the column (Fig. 2, lane 6) displayed a significant enrichment for several 21–28K (M_r 21,000–28,000) proteins. N-terminal sequence analysis of the electroblotted proteins⁷ in the 21–28K region of the gel revealed a single sequence, NH₂-DRHTVF-W[N]SSNPKFRNEDYTIHVQ (single-letter amino-acid code; the yield of cycle 8 (N) was diminished, indicating glycosylation at this predicted N-linked glycosylation site). A computer based homology search of the NBRF protein database⁸ resulted in the unambiguous assignment of this sequence to B61⁴.

Subsequent cloning and expression of full-length human B61 cDNA in CHO cells confirmed that the product of this gene has ECK-X binding activity. Culture supernatants from these transfected cells displayed ECK-X binding activity on the BIAcore, and recombinant human B61 (rhB61), purified by immobilized ECK-X receptor affinity chromatography, or by conventional chromatographic techniques, was also active. Culture supernatants from untransfected CHO cells or CHO cells expressing B61 containing an internal deletion displayed no detectable receptor binding activity.

Crosslinking experiments confirmed that rhB61 was able to bind the full-length, cell-associated ECK receptor. ¹²⁵I-rhB61 could be crosslinked to CHO cells expressing ECK (Fig. 3, lane 1) but not to parental cells (Fig. 3, lane 2). Chemical crosslinking resulted in the detection of a 145K protein-complex, as well as higher M_r forms, which could be immunoprecipitated by antibodies specific for ECK (data not shown). Unlabelled rhB61 at 50-fold molar excess (1 μ g ml⁻¹) to the labelled protein competed

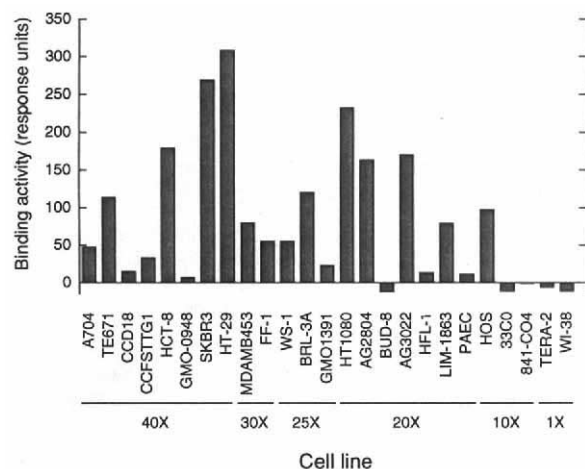


FIG. 1 Screening of cell culture supernatants for binding activity to immobilized ECK-X. Cell culture supernatants, produced under serum-depleted conditions, were concentrated 10–40 times with 10,000 MWCO membranes (Centricon 10, Amicon, Danvers, MA) and injected (30 μ l, 5 μ l min⁻¹) onto ECK-X surfaces on the BIAcore (Pharmacia Biosensor, Piscataway, NJ). Receptor binding was measured as the response at 25 s post-injection, relative to a baseline determined immediately before injection. HCT-8 and SK-BR-3 cell supernatants were selected for fractionation. Production of ECK-X binding activity by HT-1080, AG2804 and AG3022 cell lines was variable.

METHODS. For BIAcore analysis, purified ECK-X (0.2 mg ml⁻¹ in 10 mM sodium acetate, pH 4.0) was immobilized³ by two consecutive 50 μ l injections. Unreacted binding sites were blocked by injection of 1 M ethanolamine-HCl, pH 8.5, followed by washing with HEPES-buffered saline, pH 7.4, with 3.4 mM EDTA and 0.05% Tween 20, until a stable baseline was achieved. Typical immobilizations resulted in the establishment of baselines 6,000–8,000 response units above original values. Regeneration was by 10- μ l injections of 50 mM Na-acetate, 500 mM NaCl, pH 4.0. Recombinant ECK-X, which corresponds to the extracellular domain of ECK truncated at Asn 534, was expressed in CHO cells using the pDSR α 2 vector⁶, and was purified from the conditioned medium from CHO cells as follows. Culture supernatant was concentrated and diafiltered against 10 mM Tris-HCl, pH 7.4 and loaded onto an anion exchange column (Hema-Q, 10 μ m particle size, Alltech, Deerfield, IL). The column was eluted with a linear gradient of 0–0.5 M NaCl in 10 mM Tris-HCl, pH 7.4. Fractions were analysed by SDS-PAGE and immunoblotting using a rabbit antiserum generated against a synthetic N-terminal peptide of ECK. Fractions containing ECK-X were pooled, dialysed against 10 mM Tris-HCl, pH 7.4, and reappplied to the Hema-Q column. The resulting pool was concentrated (Centriprep-10, Amicon, Danvers, MA) and applied to a Superdex 200 (Pharmacia) gel filtration column equilibrated in PBS. A pool containing the purified ECK-X was made and served as the basis of further experiments.

for binding and crosslinking (Fig. 3, lane 4). Recombinant fibroblast growth factor (rFGF) at 1 μ g ml⁻¹ had no effect on rhB61 crosslinking (data not shown). Experiments using the LIM 2405 human colon carcinoma cell line⁹ yielded similar results.

Scatchard analysis¹⁰ of the steady-state binding of ¹²⁵I-rhB61 to LIM 2405 cells revealed a single-affinity class of B61 surface receptors with a dissociation constant (K_d) of 28 ± 3 nM at a density of 1.3×10^6 receptors per cell. Analysis¹¹ of the kinetic parameters of rhB61 binding to immobilized ECK-X surfaces resulted in an estimated K_d of 24 ± 4 nM.

The known polypeptide ligands for receptor protein-tyrosine kinases bind to their cognate receptors at the cell surface and stimulate tyrosine phosphorylation. To determine if B61 stimulates ECK phosphorylation, LIM 2405 cells were examined because of their high levels of ECK and low levels of endogenous ECK phosphorylation. Immunoblot analysis of ECK immuno-

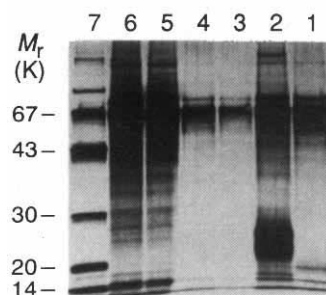
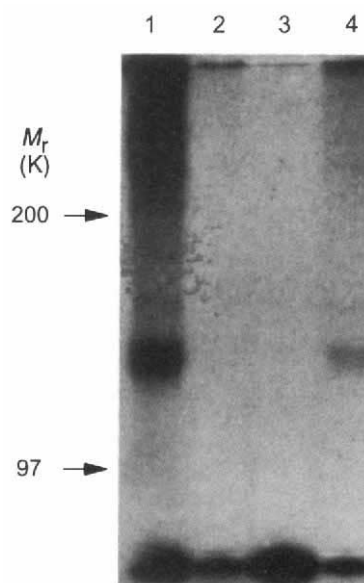


FIG. 2 SDS-PAGE analysis of immobilized ECK-X affinity chromatography. Conditioned medium (2 l) from HCT-8 cells was concentrated 20–40 times, diafiltered against PBS and loaded onto a 1 ml column of immobilized ECK-X (1 mg ECK-X per ml of CNBr-activated Sepharose; Pharmacia). Samples were concentrated and analysed by SDS-PAGE in the presence of reducing agent on 12% gels. Gels were stained with silver¹⁶. Lane 1, molecular size standards; lane 2, column load; lane 3, unbound fraction; lane 4, PBS wash; lanes 5, 6, 7, fractions from pH 4.0 elution. Lane 6, the main elution peak. The highly enriched 21–28K proteins in lane 6 were identified as B61. These proteins were not detected when a similar volume of unconditioned (mock) medium was passed over the same column, indicating that the B61 was produced by HCT-8 cells. In contrast, the higher M_r proteins present in the pH 4.0 elution fractions were also observed in control runs of mock medium. These proteins represent nonspecific interactions with the column, and originate from the serum used for cell culture.

FIG. 3 Chemical crosslinking of ¹²⁵I-rhB61 to CHO cells expressing ECK. Lane 1, ¹²⁵I-rhB61 + CHO 19.32 (expressing ECK); lane 2, ¹²⁵I-rhB61 + CHOd (untransfected); lane 3, ¹²⁵I-rhB61 + CHO 19.32, no crosslinker added; lane 4, ¹²⁵I-rhB61 + CHO 19.32 + 1 μ g unlabelled rhB61.

METHODS. CHO cells stably expressing a cDNA encoding the full-length ECK receptor (CHO 19.32), or untransfected parental cells (CHOd-), were grown to a density of 5×10^5 cells ml⁻¹, pelleted, and resuspended in PBS. For each crosslinking reaction, 2×10^6 cells were mixed with 20 ng ¹²⁵I-rhB61 in a total volume of 1 ml. The mixture was incubated at 4 °C for 1 h, followed by addition of 20 μ l of 10 mM disuccinimidyl suberate crosslinker and incubation at room temperature for 30 min. Cells were washed three times in PBS, then lysed in 0.5% NP-40, 1 mM EDTA, 1 mM phenylmethylsulphonyl fluoride in PBS. After centrifugation, the soluble fraction of each lysate was analysed by SDS-PAGE. For labelling, 5 or 10 μ g of rhB61 in 0.1 M sodium phosphate (NaPO₄, pH 8.0) was added to 5 mCi of dried ¹²⁵I-Bolton-Hunter reagent (NEN, Boston, MA) in a final volume of 50 or 100 μ l, and the tube was incubated at 4 °C for 1 h. The reaction was terminated by addition of 0.3 ml of 0.2 M glycine in 0.1 M NaPO₄, followed by incubation at 4 °C for 5 min. Labelled protein was separated from unincorporated reagent on Sephadex G-25 M (Pharmacia) equilibrated in 0.1 M NaPO₄, 0.25% gelatin. Specific activity of the ¹²⁵I-rhB61 ranged from 4 to 19 c.p.m. pg⁻¹.



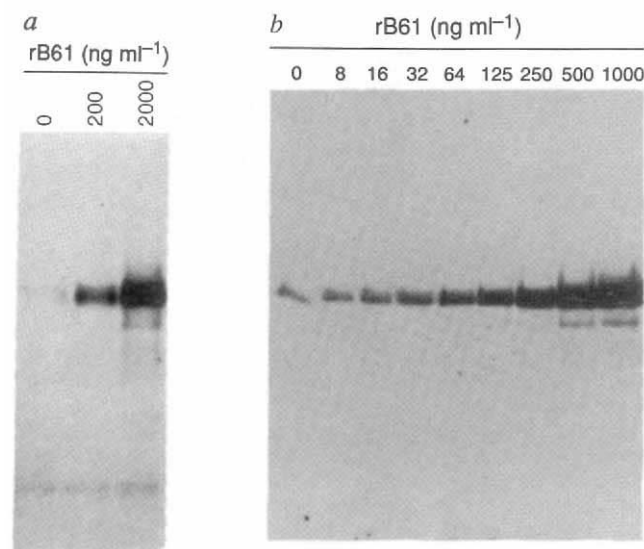


FIG. 4 rhB61 activation of the ECK receptor protein-tyrosine kinase in LIM 2405 cells. *a*, Lysates from rhB61-treated cells were immunoprecipitated with anti-ECK antibody¹, resolved by SDS-PAGE, electroblotted, and probed with anti-phosphotyrosine antibody (4G10, UBI, Lake Placid, NY). *b*, Lysates were immunoprecipitated with anti-phosphotyrosine antibody. Resulting blots were probed with anti-ECK antibody.

METHODS. Serum-starved LIM 2405 cells were treated for 20 min at 37 °C with purified rhB61, then lysed in: *a*, RIPA buffer or *b*, NP-40 buffer¹⁷. Lysates were clarified by centrifugation and incubated with the immunoprecipitating antibody for 1 h on ice. Immune complexes were adsorbed to protein-G Sepharose beads (Pharmacia) at 4 °C for 30 min, washed with lysis buffer three times, and solubilized in SDS-PAGE sample buffer. Immunoprecipitates were resolved on 7.5% gels, electrophoretically transferred to Immobilon-P (Millipore, Bedford, MA) and probed with antibodies. Immune complexes were detected by horseradish peroxidase conjugated secondary reagents using chemiluminescence (ECL, Amersham).

precipitates with anti-phosphotyrosine antibodies showed that rhB61 stimulated ECK tyrosine phosphorylation in these cells (Fig. 4*a*). This rhB61 response was dose-dependent, and became detectable at 16–32 ng ml⁻¹ (Fig. 4*b*). Phosphoamino acid analysis¹² of ECK immunoprecipitated from untreated or rhB61 treated (10 ng ml⁻¹ and 100 ng ml⁻¹) serum-starved LIM 2405 cells labelled with ³²P-orthophosphate confirmed that rhB61 induced phosphorylation on tyrosine residues, as well as on serine residues (data not shown). Treatment of LIM 2405 cells with rhB61 did not result in spurious phosphorylation of EGF receptor, nor did EGF treatment induce ECK phosphorylation (data not shown).

The concentration of B61 (16–32 ng ml⁻¹) required to induce tyrosine phosphorylation of ECK is higher, and the affinity of rhB61 for the receptor ($K_d \approx 20$ –30 nM) is somewhat lower than the equivalent values observed for many, but not all growth factors. One explanation is that B61 functions *in vivo* primarily as a membrane-anchored form, because the B61 sequence suggests that it may associate with the membrane through a glucosyl phosphatidylinositol (GPI) anchor⁴. Indeed, we have observed membrane-associated forms of rhB61, expressed in CHO cells, which can be released by phospholipase treatment. CHO cells expressing surface forms of rhB61, exhaustively washed to remove soluble rhB61, are capable of inducing tyrosine phosphorylation of ECK when co-incubated for short times with ECK-expressing cells (data not shown).

Another possibility is that B61 is in fact a ligand for another member of the EPH/ECK receptor family. Although this cannot be excluded, we have examined the binding of B61 to soluble forms of two other EPH/ECK family receptors and, although the soluble form of a human homologue of Sek¹³ (HEK8;

G.M.F. *et al.*, manuscript in preparation) inhibited binding of rhB61 to immobilized ECK-X, it was a 10-fold less effective competitor, on a molar basis, than ECK-X. Similarly, a soluble form of HEK^{13,14} inhibited binding of rhB61 to immobilized ECK-X, but with less than half the effectiveness of HEK8. rhB61 did not induce phosphorylation of several other EPH/ECK family receptor protein-tyrosine kinases (data not shown). In addition, the tissue distribution of B61 messenger RNA more closely matches that of ECK than other EPH/ECK family members (G.M.F. *et al.*, manuscript in preparation). Taken together, these data suggest that B61 is a ligand for the ECK receptor.

We have not observed a mitogenic response when ECK-expressing cells are treated with rhB61. Because binding of ligand to a chimaeric receptor containing the cytoplasmic domain of Elk, another member of the EPH/ECK family, does not induce mitogenesis¹⁵, there is no compelling reason to believe that ECK acts in this manner. Treatment of ECK-expressing cells with B61 does induce distinct changes in cellular phosphotyrosine levels, and we are studying which proteins are involved, and what cellular responses are elicited. The availability of an ECK ligand should enable us to determine what cellular responses are elicited by activation of ECK.

By analogy with other families of receptor protein-tyrosine kinases, such as the FGF and NGF receptor families, which interact with members of ligand families, one might predict that B61 is a member of a family of structurally related proteins that are ligands for EPH/ECK family receptor protein-tyrosine kinases. On the basis of the large number of receptors in this family, we anticipate that additional B61-related ligands will be identified. □

Received 17 January; accepted 23 February 1994.

1. Lindberg, R. A. & Hunter, T. *Molec. cell. Biol.* **10**, 6316–6324 (1990).
2. Fägerstam, L. *Tech. Protein Chem.* **2**, 65–71 (1991).
3. Johnsson, B., Löfås, S. & Lindquist, G. *Analyt. Biochem.* **198**, 268–277 (1991).
4. Holzman, L. B., Marks, R. M. & Dixit, V. M. *Molec. cell. Biol.* **10**, 5830–5838 (1990).
5. Sajjadi, F. G. & Pasquale, E. B. *Oncogene* **8**, 1807–1813 (1993).
6. DeClerck, Y. A., Yean, T.-D., Lu, H. S., Ting, J. & Langley, K. L. *J. biol. Chem.* **266**, 3893–3899 (1991).
7. Fausset, P. R. & Lu, H. S. *Electrophoresis* **12**, 22–27 (1991).
8. Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
9. Whitehead, R. H., Zhang, H. H. & Hayward, I. P. *Immun. Cell Biol.* **70**, 227–236 (1992).

10. Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660–672 (1949).
11. Zhou, M. *et al. Biochemistry* **32**, 8193–8198 (1993).
12. Boyle, W. J., van der Geer, P. & Hunter, T. *Meth. Enzym.* **201**, 110–148 (1991).
13. Nieto, M. A., Gilardi-Hebenstreit, P., Charnay, P. & Wilkinson, D. G. *Development* **116**, 1137–1150 (1992).
14. Boyd, A. W. *et al. J. biol. Chem.* **267**, 3262–3267 (1992).
15. Lhoták, V. & Pawson, T. *Molec. cell. Biol.* **13**, 7071–7079 (1993).
16. Merrill, C. R. in *Meth. Enzym.* **182** 477–487 (1990).
17. van der Geer, P. & Hunter, T. *EMBO J.* **12**, 5161–5172 (1993).

ACKNOWLEDGEMENT. We thank P. Steffner for his contributions.

DNA-like double helix formed by peptide nucleic acid

Pernilla Wittung*, Peter E. Nielsen†‡, Ole Buchardt‡§, Michael Egholm§ & Bengt Nordén*†

* Department of Physical Chemistry, Chalmers University of Technology, S-41296 Gothenburg, Sweden

† Research Center for Medical Biotechnology, The Panum Institute, Department of Biochemistry B, Blegdamsvej 3c, 2200 N Copenhagen, Denmark

‡ Department of Organic Chemistry, The H.C. Ørsted Institute, Universitetsparken 5, 2100 Ø Copenhagen, Denmark

ALTHOUGH the importance of the nucleobases in the DNA double helix is well understood, the evolutionary significance of the deoxyribose phosphate backbone and the contribution of this chemical entity to the overall helical structure and stability of the double helix is not so clear. Peptide nucleic acid (PNA)¹⁻⁷ is a DNA analogue with a backbone consisting of *N*-(2-aminoethyl)glycine units (Fig. 1) which has been shown to mimic DNA in forming Watson-Crick complementary duplexes with normal DNA⁷. Using circular dichroism spectroscopy we show here that two complementary PNA strands can hybridize to one another to form a helical duplex. There is a seeding of preferred chirality which is induced by the presence of an L- (or D-) lysine residue attached at the carboxy terminus of the PNA strand. These results indicate that a (deoxy)ribose phosphate backbone is not an essential requirement for the formation of double helical DNA-like structures in solution.

To investigate whether formation of a PNA-PNA duplex was possible, we designed two complementary PNA decamers that included all four nucleobases without being self-complementary: H-GTAGATCACT-L-Lys-NH₂ and H-AGTGATCTAC-L-Lys-NH₂, where H is a free amino group and L-Lys-NH₂ is a terminal lysine carboxamide. The chirality of the L-lysine amide should in principle influence the handedness of any potential helical structure formation.

The circular dichroism (CD) in the nucleobase absorption region of DNA arises from the helical stacking of the bases as a result of exciton interactions between the transitions in neighbouring bases and from their interactions with transitions of the chiral deoxyribose moiety^{8,9}. By contrast, the PNA backbone is achiral and the electronic interaction between the majority of the bases and the chiral C-terminal lysine is small. Thus any CD spectrum would be attributable to a chiral orientation of the base pairs relative to each other. The individual PNA decamers give rise to a very small CD (Fig. 2), indicating the absence of any preferred helical stacking of the bases. However, upon titration of one decamer with the complementary decamer, an intense CD spectrum arises, reaching a maximum very close to a 1:1 stoichiometric ratio of the decamers (Fig. 2, inset). The CD spectrum of the PNA-PNA complex resembles that of a DNA decamer duplex with the same sequence. When the C-terminal lysine was replaced with phenylalanine or isoleucine, the CD spectra were essentially the same, suggesting that the amino acid only plays the role of a chiral bias for the choice of helicity of a stack of bases. Substitution of D-lysine gave rise to the mirror-image spectrum, as expected, showing that the structure has now acquired opposite chirality (Fig. 2). The amplitude of the CD spectrum is comparable to that of the DNA-DNA double helix. This indicates that a substantial fraction of the PNA-PNA duplexes adopt double helices of the same handedness.

The thermal stability was found to be considerably higher for the PNA-PNA duplex (melting temperature $T_m = 67^\circ\text{C}$) than

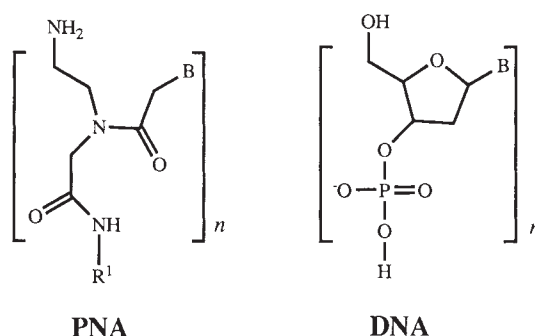


FIG. 1 Chemical structure of PNA and DNA, where B is the nucleobase, and R^1 is either a hydrogen atom or lysine amide for the PNAs investigated here.

for the DNA-PNA (antiparallel) duplex ($T_m = 51^\circ\text{C}$) or the DNA-DNA duplex ($T_m = 33.5^\circ\text{C}$) (data not shown). The thermal stability of the PNA duplex was not affected by the absence or presence of D/L-lysine amide on either strand. However, the thermal stability of the antiparallel PNA duplex was $\sim 20^\circ\text{C}$ more stable than the corresponding parallel PNA duplex (Table 1), a behaviour analogous to that of the PNA-DNA duplexes⁷. Introduction of non Watson-Crick base pairs (such as T·T or C·C) into the PNA duplex resulted in a significant decrease in T_m , indicating that the normal PNA duplex is Watson-Crick base-paired. The PNA duplex must be a double-helical structure with the helicity determined by the terminal lysine and involving a considerable number of base pairs, as indicated by the observation that the C·C mismatch at the fourth base from the lysine substantially alters (reduces) the CD spectrum (results not shown).

The development of ellipticity at 220 nm as a function of the time (Fig. 3 inset) indicates that the helix formation between the two complementary PNA oligomers is relatively slow compared with DNA-DNA duplex formation¹⁰. This time dependence of the CD may be fitted to first-order kinetics, and by measuring the CD change as a function of temperature we could determine the activation parameters (Fig. 3). Repeating these experiments using different initial PNA concentrations gave identical results, and hypochromicity studies showed an immediate (< seconds) decrease of absorption (data not shown), indicating that the PNA-PNA association itself is fast within the timescale of the experiments. Thus, the time-dependent CD must be attributed

TABLE 1 Thermal stabilities of PNA-PNA duplexes

PNA-1	PNA-2	T_m
A-L-Lys-NH ₂	B-L-Lys-NH ₂	67.0 °C
A-L-Lys-NH ₂	B-D-Lys-NH ₂	69.0 °C
A-L-Lys-NH ₂	B-NH ₂	67.5 °C
A-D-Lys-NH ₂	B-NH ₂	67.0 °C
A-NH ₂	B-NH ₂	67.0 °C
A-L-Lys-NH ₂	C-L-Lys-NH ₂	45.5 °C
D-L-Lys-NH ₂	B-L-Lys-NH ₂	51.0 °C
E-L-Lys-NH ₂	B-L-Lys-NH ₂	46.5 °C
E-L-Lys-NH ₂	F-NH ₂	$\leq 42^\circ\text{C}$

Sequences: A: H-GTAGATCACT-; B: H-AGTGATCTAC-; C: H-CATCTA GTGA-; D: H-GTAGTTCAC-; E: H-GTACATCACT-; F: H-ACTGATCTAC-; The PNAs are written from the amino to the carboxy terminal. 'H' signifies a free amino group, 'NH₂' signifies a terminal carboxamide. PNAs were synthesized as previously described²⁻⁴. They were purified by HPLC and their composition was verified by mass spectrometry. Absorbance versus temperature was measured at 260 nm in 100 mM NaCl, 10 mM sodium phosphate, 0.1 mM EDTA, pH 7. The T_m is the temperature at which the first-order derivative of the melting curve shows a maximum. All melting curves exhibited monophasic transition. PNA concentrations were 3 μM .

‡ To whom correspondence should be addressed.

FIG. 2 Circular dichroism spectra recorded at 1-cm pathlength of the PNA-PNA duplex with L-lysines (spectrum 1), the PNA-PNA duplex with D-lysines (spectrum 2), and a DNA-DNA duplex of the same sequence (spectrum 3). The CD spectrum for the PNA-PNA duplex without terminal lysines follows the baseline. The concentration was 50 μ M in base pairs. Spectra of the two single-stranded PNAs H-GTAG-ATCACT-L-Lys-NH₂ (spectrum A) and H-AGTGATCTAC-L-Lys-NH₂ (spectrum B) are also shown (50 μ M in nucleobases). In the insert, saturation is shown for the titration of PNA H-GTAGATCACT-L-Lys-NH₂ (A) and H-AGTGATCTAC-L-Lys-NH₂ (B) at equimolar amounts of the two decamers. Mass spectrometry analysis confirmed the formation of a PNA duplex.

METHODS. Hybridizations were done in a 5 mM sodium phosphate buffer, pH 7.0, at 20 °C, and the spectra were recorded after 20 min incubation. Extinction coefficients for the bases in DNA were used.

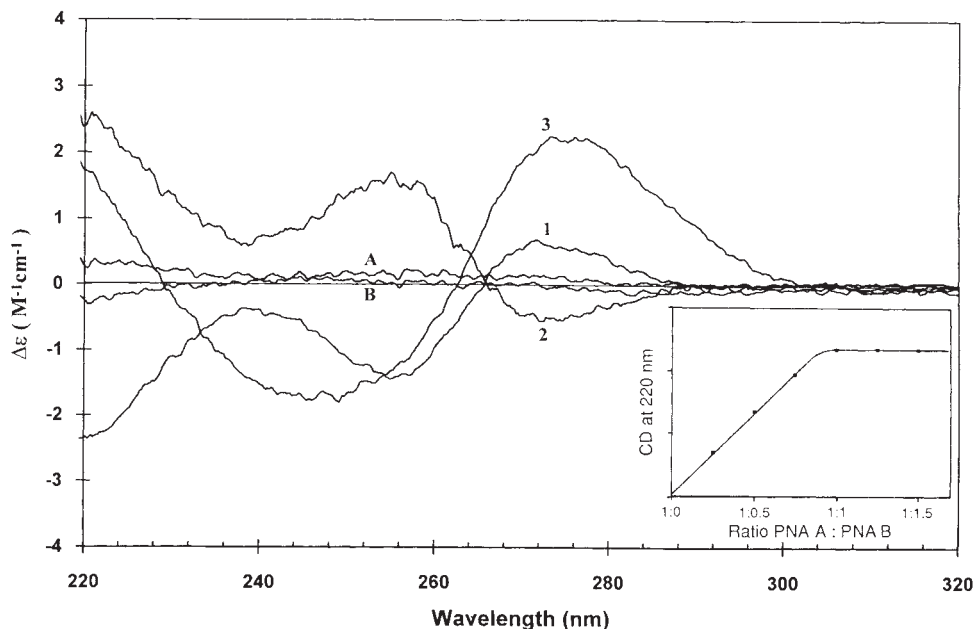
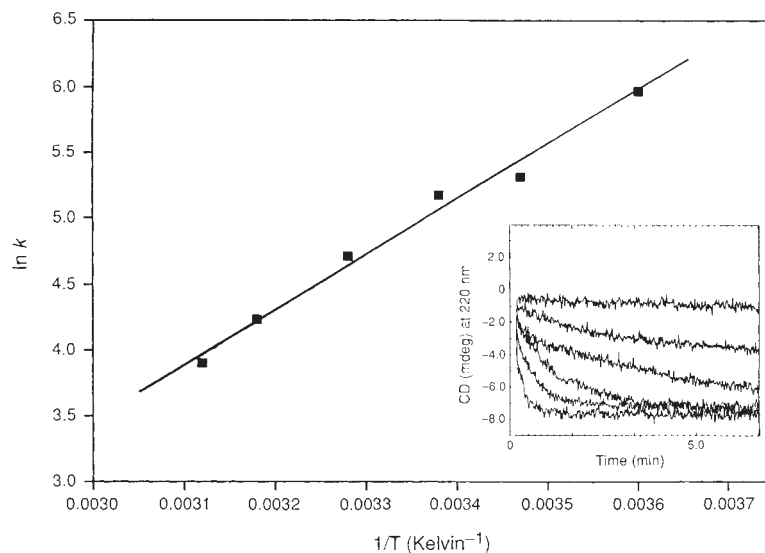


FIG. 3 Inset, development of negative circular dichroism (at 220 nm) as a function of time after mixing equimolar amounts of A with B (Fig. 2). From top to bottom, curves were recorded at 5 °C, 15 °C, 23 °C, 32 °C, 41 °C and 47 °C. Otherwise conditions were as described for Fig. 2. Main figure, Arrhenius plot for the rates from the CD kinetics. The plot provides the activation enthalpy $\Delta H^\ddagger = 33.9 \text{ kJ mol}^{-1}$ (with the approximation that $(k_B T/h) \exp(\Delta S^\ddagger/R)$ is constant, where k_B is the Boltzmann constant and h the Planck constant). The full rate equation $k = (k_B T/h) \exp(-\Delta H^\ddagger/RT) \exp(\Delta S^\ddagger/R)$ gives $\Delta S^\ddagger = -173 \text{ J (K mol)}^{-1}$.



to a reorganization process in an already base-paired complex. In the corresponding DNA-DNA and DNA-PNA decamer duplexes there is no such reorganization because the presence of the local chirality, due to the deoxyribose residues in DNA, determines immediately the handedness of base stacking at the position of each base pair in the duplex. The essentially identical melting temperatures of PNA-PNA duplexes with and without chiral centres strongly indicate that their structures are similar (helical). But in the absence of a chiral terminal amino acid, a racemic mixture of double helices of opposite helicity is formed (explaining why there is no CD).

We conclude that the association of the two PNA oligomers consist of a fast base-pairing step resulting in a 1:1 racemic mixture of right- and left-handed double helices. This step is followed by a relatively slow inversion of the double helix to the preferred chirality by the terminal lysine amide residues. The activation energy for the latter process is $\sim 34 \text{ kJ mol}^{-1}$, and the entropy of activation is large and negative, $\Delta S^\ddagger = -173 \text{ J (K mol)}^{-1}$ (Fig. 3), as expected for the highly ordered transition state that would result from changing the total helicity of the

double helix. The complete process may be described as a cooperative seeding of chirality, beginning from the terminal base pair and migrating through the stack of bases.

The implications of our results are threefold. First, considering the structural stability of the genetic material, we find that the deoxyribose backbone is not essential for the formation of helical DNA-like duplexes, which is an important result in view of the fact that an artificial homo-DNA having hexose instead of ribose in the backbone forms a parallel side-by-side duplex rather than a double helix¹¹. Second, PNA provides insight into the helical base-stacking of DNA: in contrast to double-stranded DNA, in which deoxyribose residues determine the local helicity, the helicity of the PNA-PNA duplex is determined by the cooperative arrangement of the nucleobases. PNA is therefore a relatively simple model that could be used to study optical activity in DNA itself. Third, from the point of view of evolution of life on Earth, a phosphate backbone may not have been an essential component of any prebiotic genetic material¹². Moreover, chirality may have been introduced in the biosphere by helical seeding of initially achiral polymers such as PNA. □

Received 3 November 1993; accepted 28 January 1994.

- Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. *Science* **254**, 1497–1500 (1991).
- Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. *J. Am. chem. Soc.* **114**, 1895–1897 (1992).
- Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. *J. Am. chem. Soc.* **114**, 9677–9678 (1992).
- Egholm, M. et al. *J. chem. Soc. chem. Commun.* 800–801 (1993).
- Cherny, D. Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 1667–1670 (1993).
- Kim, S. K. et al. *J. Am. chem. Soc.* **115**, 6477–6481 (1993).
- Egholm, M. et al. *Nature* **365**, 566–568 (1993).
- Cantor, C. R. & Schimmel, R. R. *Biophysical Chemistry* 4th edn (Freeman, New York, 1980).
- Norden, B., Kubista, M. & Kurucsev, T. Q. *Rev. Biophys.* **25**, 51–170 (1992).
- Tibanyend, N. et al. *Eur. J. Biochem.* **139**, 19–27 (1984).
- Eschenmoser, A. & Loewenthal, E. *Chem. Soc. Rev.* **21**, 1–16 (1992).
- Nielsen, P. E. *Origins of Life* **23**, 323–327 (1993).

Expression cloning of a mammalian proton-coupled oligopeptide transporter

You-Jun Fei^{*†}, Yoshikatsu Kanai^{*‡},
Stephan Nussberger^{*}, Vadivel Ganapathy[†],
Frederick H. Leibach[†], Michael F. Romero[§],
Satish K. Singh[§], Walter F. Boron[§]
& Matthias A. Hediger^{*||}

^{*} Renal Division, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115, USA

[†] Department of Biochemistry and Molecular Biology, Medical College of Georgia, Augusta, Georgia 30912, USA

[§] Department of Cellular and Molecular Physiology, Yale University, School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

IN mammals, active transport of organic solutes across plasma membranes was thought to be primarily driven by the Na^+ gradient^{1–3}. Here we report the cloning and functional characterization of a H^+ -coupled transporter of oligopeptides and peptide-derived antibiotics from rabbit small intestine. This new protein, named PepT1, displays an unusually broad substrate specificity. PepT1-mediated uptake is electrogenic, independent of extracellular Na^+ , K^+ and Cl^- , and of membrane potential. PepT1 messenger RNA was found in intestine, kidney and liver and in small amounts in brain. In the intestine, the PepT1 pathway constitutes a major mechanism for absorption of the products of protein digestion. To our knowledge, the PepT1 primary structure is the first reported for a proton-coupled organic solute transporter in vertebrates and represents an interesting evolutionary link between prokaryotic H^+ -coupled and vertebrate Na^+ -coupled transporters of organic solutes.

A complementary DNA coding for a 707-amino-acid peptide transporter (PepT1) (Fig. 1a) was isolated by expression cloning using *Xenopus laevis* oocytes to screen a rabbit intestinal cDNA library for the uptake of ^{14}C -labelled dipeptide glycyl-sarcosine (Gly-Sar). When expressed in oocytes, PepT1 increased the uptake of ^{14}C -labelled Gly-Sar (100 μM) 63-fold above that of control oocytes (Fig. 2a).

The amino-acid sequence of PepT1 shows weak similarity to the H^+ -dependent nitrate transporter CHL1 from plants and a peptide permease from *Saccharomyces cerevisiae* (~25% identity)^{4,5}. There is no homology of PepT1 with any other reported sequence, including the well characterized prokaryotic H^+ -coupled transporter lactose permease⁶.

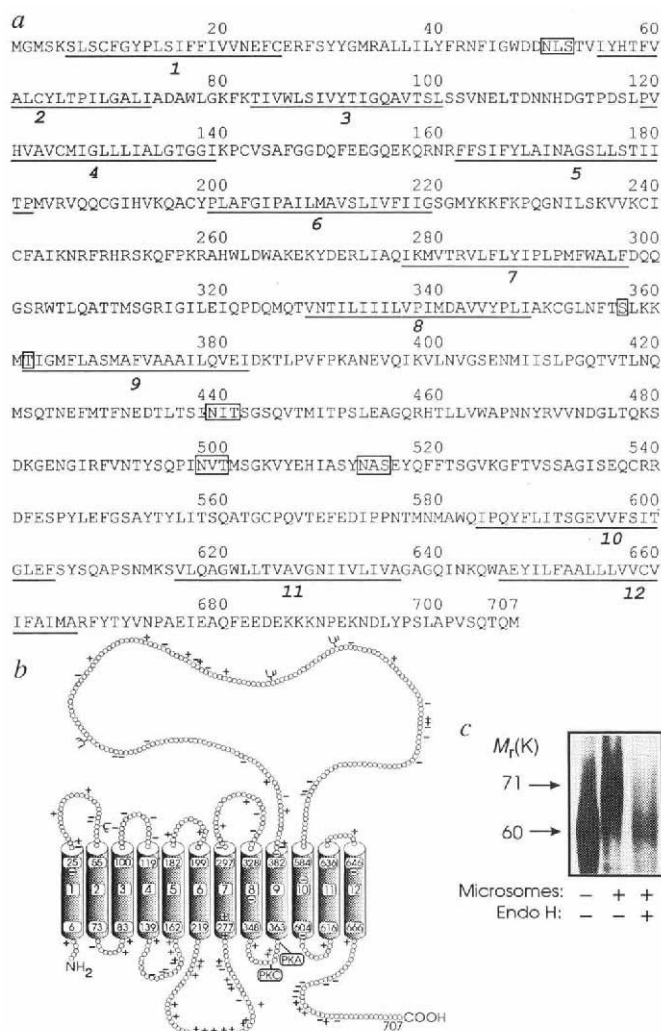


FIG. 1 a, Amino-acid sequence of PepT1. A peptide transporter cDNA (2,746 base pairs (bp)) coding for PepT1 was isolated by screening a size-selected rabbit intestinal cDNA library (2.4–4.4 kb) for the uptake of ^{14}C -labelled Gly-Sar (100 μM). The cDNA has a single consensus polyadenylation site (AATAAA) 15 nucleotides upstream from the poly(A) tail and an open reading frame from nucleotides 31 to 2,151 (not shown) that codes for a 707-amino-acid protein (PepT1). Membrane-spanning regions were predicted as described^{19,25,28} and are indicated by lines numbered 1–12. Putative N-glycosylation sites (Asn 50, 439, 498 and 513), a protein kinase C site (PKC, Ser 357) and a cAMP-dependent phosphorylation site (PKA, Ser 362) are boxed. The nucleotide sequence of PepT1 has been deposited in the GenBank database (accession no. U06467). b, Membrane model of PepT1. Putative membrane-spanning regions are depicted as cylinders. Plus and minus signs represent positively and negatively charged amino acids. Putative N-glycosylation sites (oligosaccharide trees) and PKC and PKA sites are marked. c, *In vitro* translation of PepT1 cRNA using rabbit reticulocyte lysates in the presence of microsomes. An autoradiograph is shown of an SDS–polyacrylamide gel (15%) used to analyse the *in vitro* translation products of PepT1 cRNA obtained in the absence of pancreatic microsomes (first lane) and in the presence of microsomes after centrifugation (second lane). The third lane shows the product obtained in the presence of microsomes after deglycosylation with endoglycosidase H. Translation of PepT1 cRNA in the presence of microsomes yielded a translation product of apparent M_r 71K. Treatment with endoglycosidase H revealed that at least 11K of the 71K molecular weight was due to N-linked glycosylation. This result is consistent with N-glycosylation at multiple sites²⁹.

METHODS. Expression cloning was performed as described^{19,25,28,30} and the cDNA library used was the same as that used to clone the high-affinity glutamate transporter EAAC1 (ref. 19). Procedures for *in vitro* translation have been described^{25,29}.

[†] Present address: Department of Pharmacology, Kyorin University School of Medicine, 6-20-2 Shinkawa, Mitaka, Tokyo 181, Japan.

^{||} To whom correspondence should be addressed.

The amino-acid sequence of PepT1 predicts a membrane protein with 12 membrane-spanning domains (Fig. 1b). The presence of an unusually large hydrophilic loop makes this protein structure distinct from those of previously reported organic solute transporters^{2,7}. On the basis of *in vitro* translation

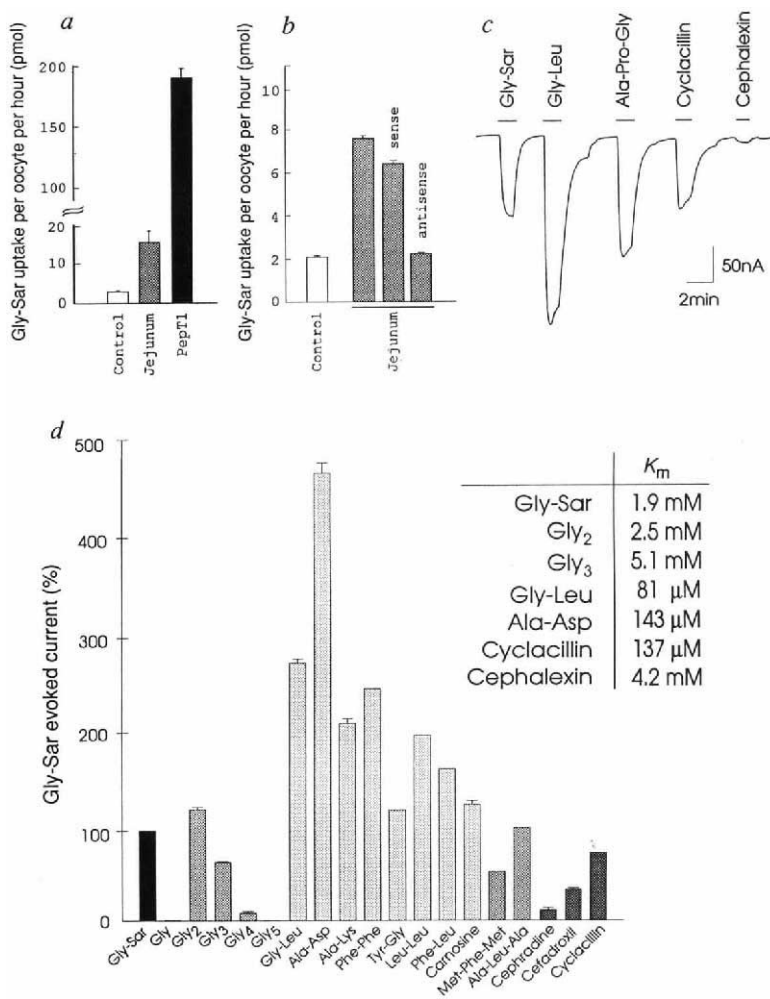
(Fig. 1c), this loop probably represents the target for *N*-linked glycosylation and is thus predicted to be extracellular.

In high-stringency northern analysis (Fig. 3), PepT1 mRNA (2.9 kilobases (kb)) was abundantly expressed in small intestine, at a lower level in kidney and liver, and only weakly in brain.

FIG. 2 Expression and functional characterization of PepT1 in *Xenopus* oocytes. **a**, Uptake studies of ¹⁴C-Gly-Sar (100 μ M) into *Xenopus* oocytes (1 h) injected with poly(A)⁺ RNA from rabbit jejunum or with cRNA synthesized by *in vitro* transcription of PepT1 cDNA. Values represent the mean $\pm$ s.e.m. ($n = 5-8$ oocytes). **b**, Hybrid depletion of rabbit small-intestine poly(A)⁺ RNA before injection into oocytes with an antisense and a sense oligonucleotide corresponding to the 5'-end coding region of the PepT1 cDNA. Hybrid depletion with the antisense oligonucleotide completely suppressed the uptake of Gly-Sar (100 μ M), whereas the sense oligonucleotide had little effect on the uptake. Values represent the mean $\pm$ s.e.m. ($n = 5-8$ oocytes). **c** and **d**, Substrate specificity of PepT1-mediated transport was studied by two-microelectrode voltage-clamp analysis of PepT1 cRNA-injected oocytes. Current responses to 1 mM bath-applied substrates measured in uptake solution (pH 5.5) at a holding potential of -60 mV are shown. Di- and tri-peptides, as well as some peptide-derived antibiotics, evoked inward currents. In **d**, current responses are presented as per cent of the current obtained for Gly-Sar. Values represent the mean $\pm$ s.e.m. ($n = 3-5$ oocytes), except for tripeptides and cyclacillin ($n = 1$). Inset in **d**, K_m values for different substrates measured at a holding potential of -60 mV.

METHODS. Isotope flux measurements: *Xenopus laevis* oocyte expression was studied as previously described^{19,24,25,28,30}. PepT1 cRNA-injected oocytes were incubated for 1 h in uptake solution (100 mM NaCl, 2 mM KCl, 1 mM MgCl₂, 1 mM CaCl₂, 3 mM HEPES, 3 mM MES, 3 mM Tris, pH 5.5) in the presence of 100 μ M ¹⁴C-Gly-Sar (specific radioactivity, 112 mCi mmol⁻¹; custom-synthesized by the Radiochemical Centre, Amersham, UK). Oocytes were rinsed with ice-cold washing solution (uptake solution adjusted to pH 7.5 with Tris) and the radioactivity of each oocyte measured by scintillation counting. Hybrid depletion: Hybrid depletion was done as described^{25,30} using an antisense oligonucleotide (23-mer) corresponding to the 5' coding region of PepT1 cDNA and a sense oligonucleotide corresponding to the same region. Rabbit jejunum poly(A)⁺ RNA (0.5 mg ml⁻¹) was incubated with the oligonucleotide (0.25 mg ml⁻¹) in the presence of 50 mM NaCl at 42 °C for 1 h and then injected into oocytes.

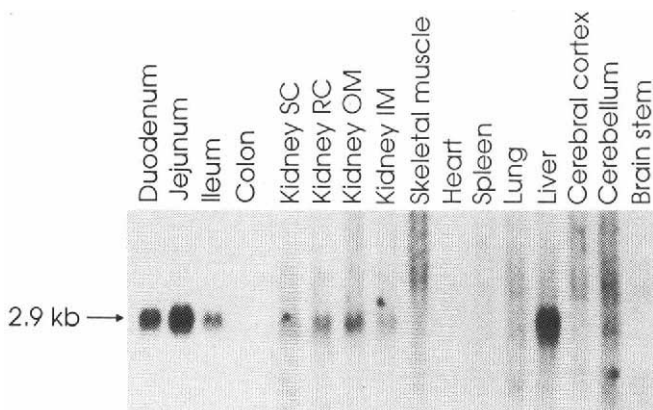
Uptakes were measured 3 days after injection. Electrophysiology: electrophysiological measurements were made 3-6 d after injection using the two-microelectrode voltage-clamp method¹⁹ (GeneClamp 500, Axon Instruments). The recording chamber was perfused with uptake solution (pH 5.5) and substrates were at times denoted by the bars (c). The



substrate specificity was determined using 1 mM bath-applied substrates. In the inset to **d**, kinetic parameters of different substrates were obtained by measuring current amplitudes at peptide concentrations ranging from 10 μ M to 10 mM; K_m values were calculated using the Eadie-Hofstee equation.

FIG. 3 High-stringency northern analysis of poly(A)⁺ RNA (3 μ g) from rabbit tissues probed with ³²P-labelled PepT1 cDNA. Left lanes (duodenum, jejunum, ileum, colon), 3-h exposure. Right lanes (all other tissues), 8-d exposure.

METHODS. Northern blotting was done as described¹⁹; hybridization was at 42 °C in 50% formamide, using ³²P-labelled full-length PepT1 cDNA as a probe. The probe was labelled using the T7 QuickPrime method (Pharmacia). The blot was finally washed in 0.1 $\times$ SSC/0.1% SDS at 65 °C. SC, superficial cortex; RC, remaining cortex; OM, outer medulla; IM, inner medulla.



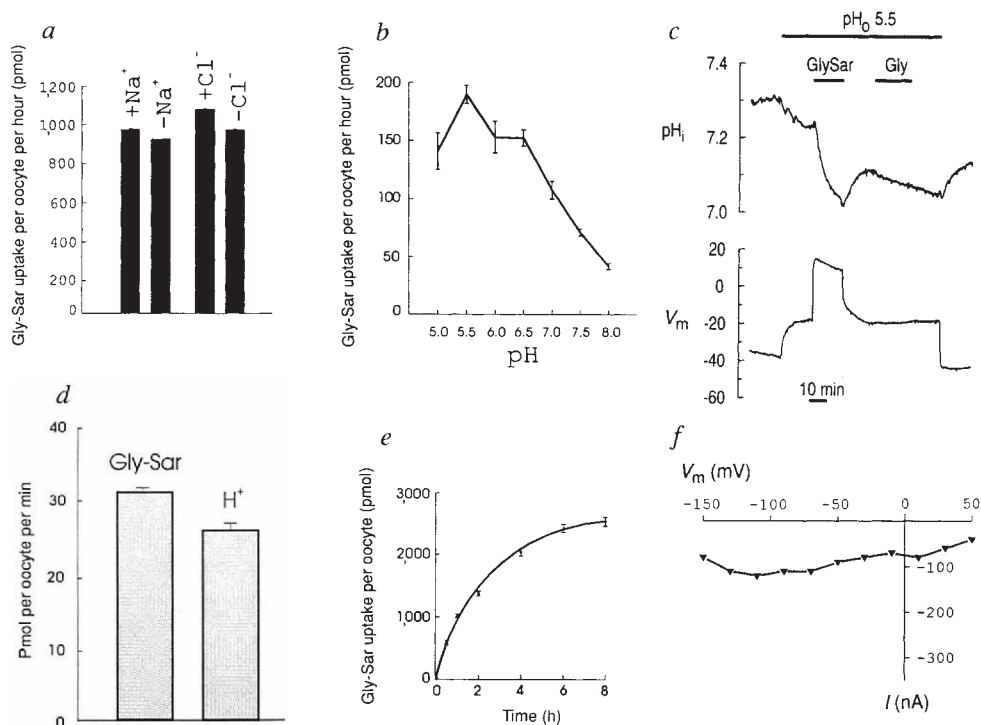
The strong signal in duodenum and jejunum is consistent with these segments being the principal site of intestinal absorption of protein digestion products³. In the kidney, peptide transporters serve to absorb filtered peptides, peptide-derived antibiotics, and peptides produced as a result of the action of luminal peptidases^{8,9}. Although an oligopeptide transporter has not previously been reported in the liver¹⁰, PepT1 in hepatocytes could be important in removing the degradation products of peptide hormones and peptide-derived drugs from the circulation¹¹. Similarly, in brain PepT1 could play a role in the clearance of degraded neuropeptides.

Previous studies have revealed that the intestine rapidly absorbs small peptides before their hydrolysis to free amino acids^{12,3}. Hybrid depletion of rabbit small-intestine mRNA (Fig. 2b) indicated that PepT1 is the only transporter responsible for transport of Gly-Sar in enterocytes. The peptide transport activities identified in intestinal apical and basolateral membranes¹³⁻¹⁵ must therefore be due to the expression of PepT1. Thus, PepT1 constitutes a major route for the absorption of protein digestion end-products.

Two-microelectrode voltage-clamp analysis of oocytes injected with PepT1 cRNA demonstrated that PepT1-mediated

FIG. 4 Stoichiometry and voltage-dependence of PepT1 transport in oocytes. Values represent the mean $\pm$ s.e.m. ($n=5-8$ oocytes) unless specified. **a**, Uptake of 14 C-labelled Gly-Sar (1 mM) in PepT1 cRNA-injected oocytes was independent of extracellular Na^+ and Cl^- . **b**, Uptake of 100 μM Gly-Sar was measured as a function of extracellular pH (pH_o). **c**, H^+ cotransport by PepT1 was demonstrated by impaling oocytes with microelectrodes to monitor intracellular pH (pH_i) and membrane potential (V_m) simultaneously. Oocytes were initially kept in pH 7.4 uptake solution. The switch to a pH 5.5 solution caused a reversible decrease in pH_i that was rapid at first, but slow near the end of the experiment. Application of Gly-Sar (10 mM) in the pH_o 5.5 solution induced a large intracellular acidification and depolarized V_m from -20 mV to $+15$ mV. In contrast, glycine, which is not a substrate of PepT1, evoked no detectable response in both pH_i and V_m . In water-injected oocytes, lowering pH_o from 7.4 to 5.5 caused a continuous pH_i decrease which was not dependent on the expression of PepT1. **d**, Comparison of the initial rate of uptake of Gly-Sar (1 mM) with the net charge flux determined using Gly-Sar-induced inward current as an indicator of H^+ flux. The H^+ to Gly-Sar flux ratio was 1:1.17. **e**, Determination of the H^+ to Gly-Sar coupling ratio based on thermodynamic analysis. Cumulative uptake of 14 C-labelled Gly-Sar (1 mM) reached a plateau in PepT1 cRNA-injected oocytes after ~ 7 h. Data represent the average of the uptakes from 20–24 oocytes after subtraction of the uptakes of water-injected oocytes. After a 10-h incubation in 1 mM non-radiolabelled 1 mM Gly-Sar, the 14 C-Gly-Sar influx was $1,754 \pm 403$ ($n=7$) pmol per oocyte in 2 h (data not shown), demonstrating that the transporter was still functional and that equilibrium had been reached. Thin-layer chromatography³⁰ of solubilized oocytes incubated in 1 mM 14 C-Gly-Sar (pH 5.5) for 7 h showed that no more than one third of the intracellular 14 C-label represented hydrolysed Gly-Sar (see below). **f**, Steady-state current-voltage (I - V) relationship of Gly-Sar-coupled H^+ inward currents in PepT1 cRNA-injected oocytes. The I - V relationship was determined in standard uptake solution (pH 5.5) under the two-electrode voltage-clamp condition. The membrane potential was held at -50 mV and stepped symmetrically to various test potentials between -150 mV and $+50$ mV for 100 ms. Gly-Sar-dependent steady-state currents were obtained as the difference in current measured in the presence and absence of 10 mM Gly-Sar. The result is representative of three experiments.

METHODS. To study the ion dependence of PepT1-mediated transport, uptake of 14 C-Gly-Sar (1 mM) was measured in standard uptake solution (pH 5.5), in Na^+ -free solution in which NaCl was replaced by choline-Cl, and in Cl^- -free solution in which Cl^- was replaced by nitrate. For intracellular pH measurement, V_m and pH_i of oocytes were measured simultaneously using microelectrodes as described^{22,23}. Electrodes



were filled with 3 M KCl and had resistances of 1–10 MΩ. For pH electrodes, silanized borosilicate pipettes were used and the tips were filled with a hydrogen ionophore (I-Cocktail B, Fluka). The rates of the Gly-Sar and H^+ influxes were determined as described²⁴. Briefly, 4 days after injection of oocytes with PepT1 cRNA, the initial rate of 14 C-Gly-Sar uptake (1 mM) was determined from the slope of the time dependence of 1–5-min uptakes at pH 5.5. Immediately after completion of the uptake measurements, the same batch of oocytes was subject to electrophysiological experiments. The membrane potential was measured in uptake solution (pH 5.5) containing 1 mM Gly-Sar. After washing with uptake solution without Gly-Sar, each oocyte was clamped at this membrane potential. Inward currents evoked by bath-applied 1 mM Gly-Sar at the determined holding potentials were recorded and the values were converted into the rate of net charge flux using Faraday's constant ($9.65 \times 10^4 \text{ C mol}^{-1}$). The equation used to calculate the electrochemical potential difference for H^+ ($\Delta\mu_{\text{H}^+}$) was $\Delta\mu_{\text{H}^+} = RT \ln ([\text{H}^+]_i / [\text{H}^+]_o) + zF(V_i - V_o)$ and that for Gly-Sar ($\Delta\mu_{\text{Gly-Sar}}$) was $\Delta\mu_{\text{Gly-Sar}} = RT \ln ([\text{Gly-Sar}]_i / [\text{Gly-Sar}]_o)$, where R is the gas constant, T the absolute temperature (295 K), F Faraday's constant, pH_i and pH_o the intra- and extracellular pH, Gly-Sar, and Gly-Sar_o, the intra- and extracellular Gly-Sar concentrations, and z the valency. Thin-layer chromatography was done as described³⁰; total extracts of oocytes in a mixture of chloroform/methanol/water were centrifuged and the supernatants spotted onto pre-coated acid-washed silica gel plates (Sigma) and chromatographed in 1-butanol/pyridine/acetic acid/water (3:2:0.6:1.5 by vol). Detection was by autoradiography or ninhydrin reaction. Unlabelled glycine, sarcosine and Gly-Sar, as well as 14 C-labelled Gly-Sar, were used as standards.

transport is electrogenic. Large inward currents were obtained when substrates such as dipeptides, tripeptides and β -lactam antibiotics (1 mM) were applied to the bath (Fig. 2c and d). In contrast, single amino acids and peptides containing more than four amino acids did not evoke a current (Fig. 2d). Transport of oligopeptide was saturable, with K_m values ranging between 137 μ M and 4.2 mM (Fig. 2d, inset). The K_m for Gly-Sar was 1.9 mM, consistent with studies of rabbit intestinal brush-border membrane vesicles ($K_m = 3.7$ mM)¹³ and *in vitro* studies of toad intestine ($K_m = 2.15$ mM)¹⁶.

The optimal length of oligopeptides preferred by PepT1 was studied using oligomers of glycine (Gly_n; $n = 1-5$) (Fig. 2d). The current was large for the glycine dimer, but decreased when n was >2 . This decrease was associated with a decrease in the affinity of the transporter (Fig. 2d, inset). Several larger peptides ($n = 5-10$) were also tested and evoked no significant current (data not shown).

Oligopeptides were transported, regardless of whether they contained acidic, basic or hydrophobic amino acids (Fig. 2). Therefore it is likely that any dipeptide can serve as a substrate of PepT1. The affinities among dipeptides, however, varied substantially (Fig. 2d, inset). PepT1 seems to have a preference for peptides containing bulky aliphatic side chains. A large inward current was evoked by the acidic dipeptide Ala-Asp (Fig. 2d). This may indicate that the β -carboxyl group of aspartic acid is transported in its protonated form.

Peptide-derived antibiotics are absorbed by means of intestinal peptide transporters^{3,17,18}. Consistent with this, the aminopenicillin cyclacillin evoked a large inward current (Fig. 2c and d). Significant currents were also observed for amino-cephalosporins such as cephalexin, cephadrine and cefadroxil. PepT1 displayed a much higher affinity for cyclacillin ($K_m = 137$ μ M) than for cephalexin ($K_m \sim 4.2$ mM).

Uptake mediated by PepT1 was completely independent of extracellular Na⁺ and Cl⁻ (Fig. 4a). Furthermore, under voltage-clamp conditions, increasing the extracellular K⁺ concentration from 2 to 50 mM¹⁹ did not suppress the current evoked by 1 mM Gly-Sar (holding potential between -120 and +30 mV; extracellular pH (pH_o) 5.5 or 7.5; data not shown), indicating that peptide transport is not coupled to the countertransport of K⁺. However, transport by PepT1 in oocytes was maximal at a pH_o of 5.5 (Fig. 4b). This is consistent with the acidic unstirred water layer (pH 5.5-6.0) in the intestinal lumen which provides the proton motive force for peptide absorption^{3,15,16,20,21}.

H⁺ cotransport of PepT1 was directly demonstrated by measuring intracellular pH (pH_i) of oocytes using a pH-sensitive microelectrode filled with a hydrogen-selective ionophore^{22,23}. Application of 10 mM Gly-Sar to PepT1 cRNA-injected oocytes caused pH_i to decrease from 7.22 to 7.0 (Fig. 4c) and the oocyte membrane to depolarize. In contrast, glycine, which is not a substrate of the transporter, evoked no detectable response. In water-injected oocytes, Gly-Sar did not affect pH_i or membrane potential (V_m) (result not shown). Thus transport of Gly-Sar not only requires a low pH_o, but also causes a decrease in pH_i. This demonstrates that PepT1 cotransports dipeptides and protons.

The coupling ratio of PepT1-mediated transport to H⁺ was examined by comparing the initial rate of Gly-Sar uptake to H⁺ movement as indicated by the inward current induced by Gly-Sar²⁴. The uptake of Gly-Sar was linear over the first 5 min. The slope of this curve was taken to represent the Gly-Sar flux rate and was 31.1 ± 0.7 pmol min⁻¹. Assuming that this current is solely carried by H⁺, the amplitude was converted to the H⁺ flux using Faraday's constant and was 26.6 ± 1.2 pmol min⁻¹. This gives a ratio of H⁺ influx to Gly-Sar uptake of 1 to 1.17 (Fig. 4d) and indicates that one H⁺ is cotransported by PepT1 with one Gly-Sar molecule.

This coupling ratio is consistent with thermodynamic analysis. At equilibrium, when the sum of the electrochemical potential differences for H⁺ ($\Delta\mu_H$) and Gly-Sar ($\Delta\mu_{\text{Gly-Sar}}$) across the

oocyte plasma membrane is zero, the theoretical intracellular Gly-Sar concentration (Gly-Sar_i) that can be attained at a given extracellular Gly-Sar concentration (Gly-Sar_o) is determined by the stoichiometry of the transporter. Given a Gly-Sar_o of 1 mM (pH_o 5.5, 22 °C), the theoretical Gly-Sar_i for a stoichiometry of 1:1 is 42 mM, whereas that for a stoichiometry of 2:1 is 1,728 mM. Uptake of ¹⁴C-labelled Gly-Sar (1 mM) reached a plateau in PepT1 cRNA-injected oocytes after ~7h (Fig. 4e). This and other evidence (Fig. 4 legend) make it reasonable to assume, as a first approximation, that the system is in thermodynamic equilibrium. After an 8-h incubation with Gly-Sar (1 mM, pH_o 5.5, 22 °C), pH_i was 6.96 ± 0.03 (mean $\pm$ s.e.m.; $n = 14$), V_m was -9.8 ± 0.8 mV, and Gly-Sar_i was $2,558 \pm 587$ ($n = 22$) pmol per oocyte (Fig. 4e). Assuming that the effective intracellular volume of oocytes is 0.25 μ l²⁵ and that one third of the dipeptide is hydrolysed (Fig. 4 legend), Gly-Sar_i is 6.8 mM. This value is much closer to the theoretical Gly-Sar_i value predicted for a stoichiometry of 1:1. Taken together, our results show that PepT1-mediated peptide transport of Gly-Sar is coupled to the cotransport of a single H⁺ ion.

Transport of Gly-Sar by PepT1 (10 mM, pH 5.5) is almost completely independent of V_m (Fig. 4f). Thus the rate-limiting step of PepT1-mediated Gly-Sar transport does not seem to involve charge movement within the transmembrane electric field. This is in contrast to other organic solute transporters such as the Na⁺/glucose cotransporter²⁶ and the GABA transporter²⁷, which show a sigmoidal current-voltage relationship.

In summary, we report the primary structure and functional characterization of PepT1, a novel proton-coupled transporter with a unique broad-substrate specificity for di-, tri- and tetrapeptides, as well as a variety of peptide-derived compounds. The identification of PepT1 mRNA in intestine, kidney, liver and brain suggests that there are common mechanisms involved in the cellular uptake of both exogenous and endogenous oligopeptides. Structural analysis of substrate-PepT1 interactions will facilitate the design of orally administered therapeutic drugs. □

Received 1 November 1993; accepted 1 February 1994.

- Crane, R. K., Miller, D. & Bihler, I. in *Membrane Transport and Metabolism* (eds Kleinzeller, A. & Kotyk, A.) 439-449 (Czech. Acad. Sci., Prague, 1961).
- Wright, E. M., Hager, K. M. & Turk, E. *Curr. Opin. Cell Biol.* **4**, 696-702 (1992).
- Ganapathy, V., Brandsch, M. & Leibach, F. H. *Physiology of the Gastrointestinal Tract* 3rd edn (ed. Johnson, L. R.) 1773-1794 (Raven, New York, 1994).
- Tsay, Y.-F., Schroeder, J., Feldmann, K. A. & Crawford, N. M. *Cell* **72**, 705-713 (1993).
- Perry, J. R., Basral, M. A., Steiner, H.-Y., Naider, F. & Becker, J. M. *Molec. cell. Biol.* **14**, 104-115 (1994).
- Kaback, H. R. *Biochim. biophys. Acta* **1101**, 210-213 (1992).
- Kanai, Y., Smith, C. P. & Hediger, M. A. *FASEB. J.* **7**, 1450-1459 (1993).
- Truppathi, C., Ganapathy, V. & Leibach, F. H. *J. Biol. Chem.* **265**, 14870-14874 (1990).
- Daniel, H. & Adibi, S. A. *J. Clin. Invest.* **92**, 2215-2223 (1993).
- Lombardo, Y. B., Morse, E. L. & Adibi, S. A. *J. Biol. Chem.* **263**, 12920-12926 (1988).
- Adedoyin, A., Perry, P. R. & Wilkinson, G. R. *Drug Metab. Dispos.* **21**, 184-188 (1993).
- Adibi, A. A. & Mercer, D. W. *J. Clin. Invest.* **52**, 1586-1594 (1973).
- Ganapathy, V., Burckhardt, G. & Leibach, F. H. *Biochim. biophys. Acta* **816**, 234-240 (1985).
- Dyer, J. et al. *Biochim. J.* **269**, 565-571 (1990).
- Thwaites, D. T., Brown, C. D. A., Hirst, B. H. & Simmons, N. L. *J. Biol. Chem.* **268**, 7640-7642 (1993).
- Abe, M., Hoshi, T. & Tajima, A. *J. Physiol.* **394**, 481-499 (1987).
- Kramer, W. et al. *Eur. J. Biochem.* **204**, 923-930 (1992).
- Sinko, P. & Amidon, G. J. *Pharmac. Sci.* **78**, 723-727 (1989).
- Kanai, Y. & Hediger, M. A. *Nature* **360**, 467-471 (1992).
- Lucas, M. *Gut* **24**, 734-739 (1983).
- Ganapathy, V. & Leibach, F. H. *J. Biol. Chem.* **258**, 14189-14192 (1983).
- Siebens, A. W. & Boron, W. F. *J. Gen. Physiol.* **90**, 799-831 (1987).
- Davis, B. A., Hogan, E. M. & Boron, W. F. *Am. J. Physiol.* **263**, C246-C256 (1992).
- Kanai, Y., Lee, W.-S., You, G., Brown, D. & Hediger, M. A. *J. Clin. Invest.* **93**, 397-404 (1994).
- You, G. et al. *Nature* **365**, 844-847 (1993).
- Parent, L., Supplisson, S., Loo, D. D. F. & Wright, E. M. *J. Membr. Biol.* **125**, 49-62 (1992).
- Mager, S. et al. *Neuron* **10**, 177-188 (1993).
- Hediger, M. A., Coady, M. J., Ikeda, T. S. & Wright, E. M. *Nature* **330**, 379-381 (1987).
- Hediger, M. A., Mendlein, J., Lee, H.-S. & Wright, E. M. *Biochim. biophys. Acta* **1064**, 360-364 (1991).
- Wells, R. G. & Hediger, M. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5596-5600 (1992).

ACKNOWLEDGEMENTS. We thank P. Boutin for technical assistance and P. R. Larsen and C. P. Smith for discussion. This work was supported by grants from the NIH (M.A.H., F.H.L. and W.F.B.) and by a long-term fellowship from the International Human Frontier Science Program Organization (Y.K.).

Enzymatic activities correlate with chimaeric substitutions at the actin-binding face of myosin

Taro Q. P. Uyeda*, Kathleen M. Ruppel & James A. Spudich

Departments of Biochemistry and of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA

MYOSINS are a functionally divergent group of mechanochemical enzymes involved in various motile activities in cells¹. Despite a high degree of conservation in the amino-acid sequence of the 130K motor domain^{2,3} (head region) of the molecule, there are large differences in the enzymatic and motile activities (Tables 1 and 2) of myosins from diverse species and cell types. However, the degree of conservation is not uniform throughout the head sequence⁴; therefore, one reasonable hypothesis is that the functional differences between myosins derive from the poorly conserved areas. The most prominent divergent region occurs at the 50K/20K junction, a region of the molecule sensitive to proteolytic digestion⁵ and a binding site for actin⁶⁻¹². We have now constructed chimaeras of this region of myosin by substituting the 9-amino-acid *Dictyostelium* junction region with those from myosins from other species and find that the actin-activated ATPase correlates well with the activity of the myosin from which the junction region was derived. Our results suggest that this region, likely to be part of the myosin head that interacts directly with actin^{10,13,14}, is important in determining the enzymatic activity of myosin.

We created chimaeric myosins containing substitutions of the 50K/20K junction region. Comparison of the primary structures of various myosins (Fig. 1) showed that the sequence homology of the 50K domain of the myosin head ends at Phe 612 in the *Dictyostelium* myosin sequence and resumes at Lys 622 in the 20K domain. Therefore, we analysed the region of sequence diversity between these residues which we define as the 50K/20K junction region.

Using this definition, we designed chimaeric *Dictyostelium mhcA* genes containing substitutions of the 50K/20K junction region. We chose the α - and β -isoforms of rat cardiac myosin, rabbit skeletal and chicken smooth muscle myosins¹⁵⁻¹⁷ as donors of the 50K/20K junction sequence. The mutated *mhcA* genes, made using the polymerase chain reaction (PCR), were named α RCM, β RCM, RSKM and CSMM, respectively. The chimaeras were expressed in a *Dictyostelium discoideum* strain lacking the single conventional myosin (myosin II; referred to here as myosin) heavy-chain (*mhcA*) gene (HS1 cells; K.M.R. *et al.*, manuscript submitted). We isolated independent cell lines expressing each of the four chimaeras and by western blot analysis showed them to contain comparable levels of myosin heavy chain to those of wild-type cells.

To assess the functionality of the chimaeric *mhcA* genes *in vivo*, we analysed the ability of the mutant cell lines to undergo cytokinesis and form fruiting bodies, processes that require functional myosin in *Dictyostelium*^{18,19}. Unlike the parent *mhcA* cell line, HS1 cells expressing any one of the four mutated *mhcA* genes recovered the ability to divide in suspension culture, with doubling times comparable to those of wild-type cells. Furthermore, cell lines expressing each of the mutated *mhcA* genes recovered the ability to form fruiting bodies (Fig. 2). These results demonstrated that, despite large variations in the length and composition of the sequences introduced into the *mhcA* gene

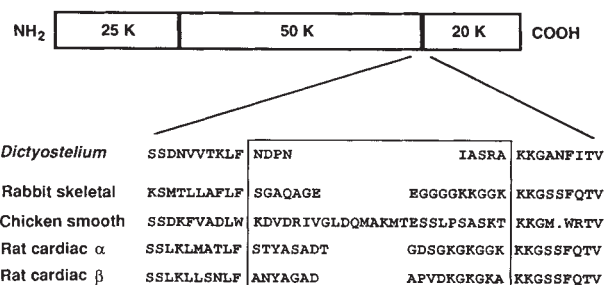


FIG. 1 Amino-acid sequences of various myosin heavy chains in a region encompassing the 50K/20K junction. The boxed sequences represent the highly divergent region defined here as the 50K/20K junction region. The wild-type sequence of the *Dictyostelium* myosin heavy-chain gene (*mhcA*) was mutated such that the boxed sequence was replaced with the corresponding sequence from the rabbit skeletal muscle, chicken gizzard smooth muscle or rat cardiac α - or β -myosin. The resultant chimaeric genes are called RSKM, CSMM, α RCM and β RCM, respectively. The CSMM also contains a point mutation of Phe 655 to Trp. Protein sequences were obtained from the Protein Identification Resource data bank (PIR), version 32, or the University of Geneva protein sequence data bank (Swiss-Prot), version 22. Accession numbers are as follows: *Dictyostelium* myosin II, PIR A26655; rabbit skeletal muscle myosin, Swiss-Prot P02562; chicken gizzard smooth muscle myosin, Swiss-Prot P10587; rat cardiac α -isoform, PIR S06005; rat cardiac β -isoform, Swiss-Prot P02564.

METHODS. The four chimaeric *Dictyostelium mhcA* genes were first made as two fragments using PCR technology; the wild-type *mhcA* sequence was used as the template. The 5' fragments were synthesized using a common 5' primer starting at nucleotide 1,510 and individual 3' mutagenic primers starting inside the region of the substitution. The 3' fragments were synthesized using a common 3' primer starting at nucleotide 2,490 and 5' mutagenic primers starting inside the substitution region. Each pair of mutagenic primers has some overlap to allow subsequent fusion of the 5' and 3' fragments either by restriction enzyme digestion followed by ligation (α RCM, β RCM, RSKM) or by fusion PCR (CSMM). The sequence between the *Bgl*II and *Acc*I sites flanking the mutated region was confirmed by dideoxy sequencing for each chimaera, and these *Bgl*II-*Acc*I fragments were used to replace the corresponding wild-type sequence of the *mhcA* gene in pBIGMyD (K.M.R. *et al.*, manuscript submitted). Sequences of the mutagenic primer pairs (5'; 3') (all sequences are shown 5' to 3'; mutated sequences are underlined): α RCM, TCAGCGCTAGCATAAGTACTGAAAAGTTTGGTGACAA-CG; CTTATGCTAGCGCTGATCTGGTGATTCGGGAAAAGGTAAGGTTGGTA-AAAAGAAAAGGTGC; β RCM, GCATCGGCGCCAGCATAATTAGCGAAAAGTT-TGGTGACAAACG; ATGCTGGCGCCGATGCTCCAGTTGATAAAGGTAAGGTA-AAGCAAAGAAAAGGTGCAAACTTTATC; RSKM, TGAGTCTCCGAGAAAAGTT-TGGTGACAAACG; AAAAGTCCGAGAGCTGAAGTGAAGAGGTTGGTGGT-TGGTAAGAAAAGGTGTAAGAAAAGGTGCAAACTTTA; CSMM, GTAAAC-TAGATTGAGTCAATTTAGCCATTGATCTAAACCAACATACGATCAACATCTT-TCCAAAGTTTGGTGACAACTTTG; TGAATCTAGTTTACCATCAGGCTTCAAAA-ACAAAAGAAAAGGTGCAAACTTT.

at the 50/20K junction region, all the chimaeric myosins are functional *in vivo*.

We then purified each myosin from the appropriate cell line, phosphorylated it using recombinant *Dictyostelium* myosin light-chain kinase (K.M.R. *et al.*, manuscript submitted), and then assayed it for the ability to hydrolyse ATP and translocate actin filaments *in vitro*. Strikingly, there is a clear correlation between the actin-activated (Mg^{2+}) ATPase activities of the chimaeric myosins and those of the myosins from which the 50K/20K junction regions were derived (Table 1). The RSKM chimaera shows the highest activity among the mutant myosins (fivefold higher than wild-type *Dictyostelium* myosin), and the CSMM chimaera shows the lowest activity (half of wild-type *Dictyostelium* myosin). The α RCM and β RCM chimaeras show intermediate activities, with the activity of the α RCM chimaera ~40% higher than that of the β RCM chimaera. The order of the activities of the chimaeric myosins exactly matches that of the donor myosins (Table 1).

* Present address: National Institute for Advanced Interdisciplinary Research, AIST, Tsukuba, Ibaraki 305, Japan.

TABLE 1 ATPase activities of the wild-type and chimaeric myosins

	High-salt Ca ²⁺ -ATPase	Mg ²⁺ -ATPase	Actin-activated Mg ²⁺ -ATPase	Actin-activated Mg ²⁺ -ATPase of parent myosin*
Wild type	2.9 ± 0.5	0.07 ± 0.05	0.78 ± 0.14	
RSKM	4.4 ± 0.9	0.12 ± 0.04	4.0 ± 0.3	20†
CSMM	4.8 ± 0.7	0.11 ± 0.01	0.44 ± 0.21	0.95‡
αRCM	4.8 ± 0.5	0.16 ± 0.01	3.2 ± 0.4	3.7§
βRCM	4.9 ± 0.3	0.19 ± 0.05	2.3 ± 0.2	1.7§

Results are given as P_i liberated per head per s. The purified wild-type and chimaeric *Dictyostelium* myosins were assayed for ATPase activity under the three conditions listed. The high-salt Ca²⁺-ATPase activity was measured in 10 mM imidazole, pH 7.4, 0.6 M KCl, 5 mM CaCl₂, 1 mM DTT and 3 mM [γ -³²P]ATP (~1 mCi mmol⁻¹). The Mg²⁺-ATPase activity was measured in 25 mM imidazole, pH 7.4, 25 mM KCl, 4 mM MgCl₂, 1 mM DTT and 3 mM [γ -³²P]ATP with or without 1 mg ml⁻¹ rabbit skeletal muscle F-actin. The concentration of myosin was 0.1 mg ml⁻¹, which was in the linear range for ATPase activity. The reaction was started by adding ATP, was allowed to proceed for 21 min at 30 °C, and the liberated P_i was quantified. Results are means ± standard deviations of three to five independent preparations and assays; these activities are compared with those of the native myosins from which the 50K/20K junction regions were donated. The purification of myosins will be described elsewhere (K.M.R. *et al.* manuscript submitted). Briefly, cells expressing the appropriate myosin were lysed in an EDTA-containing low-salt buffer. Myosin was extracted from the insoluble fraction with a Mg²⁺-ATP-containing high-salt buffer and subjected to two rounds of assembly-disassembly of myosin filaments. The resultant pure myosin was then treated with recombinant *Dictyostelium* myosin light-chain kinase before further characterization.

* 'Parent' myosin is that from which the 50K/20K junction was derived. Care must be taken when comparing the activities of different myosins because they were assayed under different conditions.

† Heavy meromyosin at 30 °C under conditions very similar to those used in the present study².

‡ Phosphorylated turkey gizzard heavy meromyosin at 25 °C (ref. 27).

§ Full-length myosins at 25 °C (ref. 15).

Analysis of the chimaeric myosins in an *in vitro* movement assay (Table 2) gave very different results. All the mutant myosins move actin filaments more slowly than the wild-type *Dictyostelium* myosin. Unlike the actin-activated (Mg²⁺)-ATPase activities, there is no correlation between the speed of the chimaeric myosins and that of the donor of the 50K/20K junction region. Therefore, the actin-activated ATPase activity and the sliding velocity are disproportionately altered by these mutations.

Previously, a comparison of the enzymatic and motile activities of skeletal muscle myosins from several animal species demonstrated a good correlation between ATPase activity and contraction speed²⁰. In terms of kinetics, however, it is not clear why these two parameters should be tightly coupled. In generally accepted models, the myosin ATPase cycle consists of alternating weak- and strong-affinity states for actin^{21,22}. The weak-affinity state is initiated by binding of ATP to the myosin head, which is then followed by hydrolysis of ATP and release of inorganic phosphate (P_i) which leads to the strong-affinity state for actin. It is believed that the force-producing portion of the cycle follows or coincides with P_i release. The rate-limiting step of this cycle is proposed to be somewhere in the weakly bound state^{22,23}. Conversely, the contraction speed should be related to the force-generating strongly bound state^{24,25} and thus be relatively unaffected by the weakly bound state. Therefore, the observed correlation between the ATPase and motile activities of the skeletal muscle myosins is probably a consequence of functional optimization of these proteins for particular purposes, rather than being causal or obligatory. It is reasonable that different structural features of the motor limit each of these activities and that mutagenesis could alter one activity without affecting the other. Similar uncoupling of the ATPase and motile activities of myosin has been observed using certain ATP analogues²⁶.

The fivefold higher actin-activated (Mg²⁺)-ATPase activity of the RSKM chimaera demonstrates that the 50K/20K junction

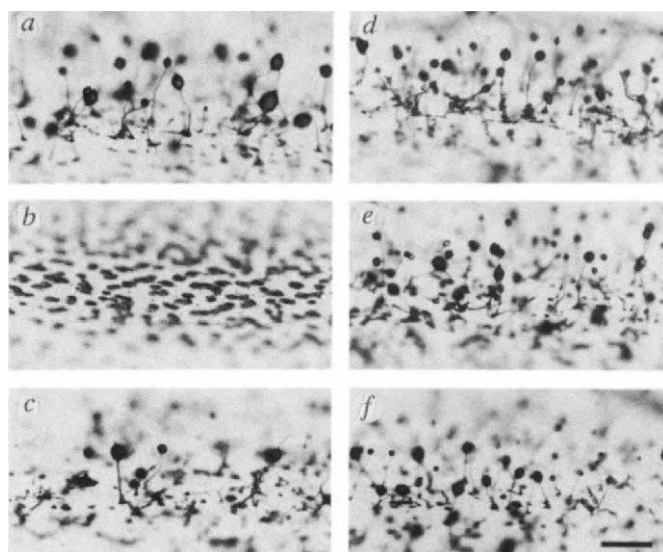


FIG. 2 Formation of fruiting bodies. a, JH10 (*mhcA*⁺); b, HS1 (*mhcA*⁻); c, RSKM; d, CSMM; e, αRCM; f, βRCM. The development of the HS1 cells is arrested at the 'mound' stage, whereas the parent JH10 cells and HS1 cells expressing any of the chimaeric *mhcA* genes can form fruiting bodies.

METHODS. The parental cell line HS1 is a G418-sensitive *mhcA*⁻ strain generated by gene replacement (K.M.R. *et al.*, manuscript submitted). The chimaeric myosins, fused to the *Dictyostelium* actin 15 promoter, were cloned into pBIG (K.M.R. *et al.*, manuscript submitted), a self-replicating *Dictyostelium* vector that carries a gene conferring resistance to the antibiotic G418. Colonies that arose in HL5 media³⁰ supplemented with 100 U ml⁻¹ penicillin, 100 μg ml⁻¹ streptomycin and 8 μg ml⁻¹ G418 (HL5/G418 media) were clonally isolated. Western blot analysis showed that independent G418-resistant cell lines expressing each of the four chimaeras contained levels of myosin heavy chain comparable to those of wild-type cells. A small number of cells from each cell line were inoculated on a lawn of *Klebsiella aerogenes* on SM/5 plates³⁰. Incubation for 3–4 days at 22 °C allowed competent cells to form fruiting bodies within 'plaques' where the food bacteria had been exhausted.

TABLE 2 Velocity of the wild-type and chimaeric myosins in an *in vitro* motility assay

	Sliding velocity (μm s ⁻¹)	Sliding velocity of parent myosin* (μm s ⁻¹)
Wild type	3.1 ± 0.3	—
RSKM	1.5 ± 0.1	7.6†
CSMM	2.2 ± 0.2	0.2‡
αRCM	1.7 ± 0.2	3§
βRCM	1.4 ± 0.2	1§

The motile activity of myosin was measured in a sliding filament *in vitro* motility assay²⁴. The phosphorylated myosin was appropriately diluted in motility assay buffer (AB; 25 mM imidazole, pH 7.4, 25 mM KCl, 4 mM MgCl₂, 1 mM EGTA, 10 mM DTT), and applied to a flow cell made of a glass slide and a nitrocellulose-coated coverslip. After blocking with AB containing 0.5 mg ml⁻¹ BSA (AB/BSA), a solution of rhodamine-phalloidin-labelled rabbit skeletal muscle F-actin in AB/BSA was introduced. Active movement was initiated by introducing AB/BSA containing 2 mM ATP, 0.8% methylcellulose and an oxygen scavenger enzyme system³. Movement at 30 °C was recorded and analysed as described elsewhere²⁴. Data are mean ± s.d. of 15 filaments.

* 'Parent' myosin is that from which the 50K/20K junction was derived. Care must be taken when comparing the activities of different myosins as they were assayed under different conditions.

† Sliding filament assay of rabbit skeletal muscle HMM, measured at 30 °C under conditions very similar to those used in the present study²⁴.

‡ Sliding filament assay of phosphorylated turkey gizzard myosin at 25 °C²⁸.

§ Sliding filament assay using rabbit cardiac α and β full-length myosin²⁹.

GUIDE TO AUTHORS

region strongly affects the rate-limiting step of the ATPase cycle of *Dictyostelium* myosin. Little is known about the nature of this rate-limiting step. In skeletal muscle S1, the hydrolysis of ATP²³ or a subsequent conformation change preceding P_i release²² is thought to be rate-limiting. It may be that the 50K/20K junction region modifies the ATPase activity of *Dictyostelium* myosin by modulating its interaction with actin. For instance, the 50K/20K junction may modulate the ATPase rate by affecting the rate of transition from the weak to the strong binding state. If this is the case, it is reasonable that the sliding velocity, which is related to the strongly bound state, does not change in accordance with the change in ATPase activity. It is not known why all four chimaeric myosins move more slowly than wild-type myosin. We are now modifying our system to express soluble fragments of the chimaeras to measure their kinetic parameters in more depth and allow us to determine more precisely how the 50K/20K junction region modifies the kinetics of the ATPase cycle. □

Received 25 October 1993; accepted 6 January 1994.

- Spudich, J. A. *Cell Regul.* **1**, 1–11 (1989).
- Toyoshima, Y. Y. et al. *Nature* **328**, 536–539 (1987).
- Kishino, A. & Yanagida, T. *Nature* **334**, 74–76 (1988).
- Warrick, H. M. & Spudich, J. A. *Rev. Cell Biol.* **3**, 379–421 (1987).
- Balint, M., Wolf, I., Tarcsafavi, A., Gergely, J. & Sreter, F. A. *Archs Biochem Biophys.* **190**, 793–799 (1978).
- Mornet, D., Pantel, P., Audemard, E. & Kassab, R. *Biochem. biophys. Res. Commun.* **89**, 925–932 (1979).
- Yamamoto, K. & Sekine, T. *J. Biochem.* **86**, 1863–1868 (1979).
- Botts, J., Muhrad, A., Takashi, R. & Morales, M. F. *Biochemistry* **21**, 6903–6905 (1982).
- Sutoh, K. *Biochemistry* **22**, 1579–1585 (1983).
- Yamamoto, K. *Biochemistry* **28**, 5573–5577 (1989).
- DasGupta, G. & Reisler, E. *J. molec. Biol.* **207**, 833–836 (1989).
- Sutoh, K., Ando, M., Sutoh, K. & Toyoshima, Y. Y. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7711–7714 (1991).
- Rayment, I. et al. *Science* **261**, 58–65 (1993).
- Schroeder, R. R. et al. *Nature* **364**, 171–174 (1993).
- Pope, B., Hoh, J. F. Y. & Weeds, A. *FEBS Lett.* **118**, 205–208 (1980).
- Schwartz, K. et al. *J. molec. cell. Cardiol.* **13**, 1071–1075 (1981).
- McNally, E. M., Kraft, R., Bravo-Zehnder, M., Taylor, D. A. & Leinwand, L. A. *J. molec. Biol.* **210**, 665–671 (1989).
- Knecht, D. A. & Loomis, W. F. *Science* **236**, 1081–1086 (1987).
- De Lozanne, A. & Spudich, J. A. *Science* **236**, 1086–1091 (1987).
- Barany, M. *J. gen. Physiol.* **50** (suppl.), 197–218 (1967).
- Lynn, R. W. & Taylor, E. W. *Biochemistry* **10**, 4617–4624 (1971).
- Stein, L. A., Schwarz, R. P. Jr, Chock, P. B. & Eisenberg, E. *Biochemistry* **18**, 3895–3909 (1979).
- Rosenfeld, S. S. & Taylor, E. W. *J. biol. Chem.* **10**, 11920–11929 (1984).
- Uyeda, T. Q. P., Kron, S. J. & Spudich, J. A. *J. molec. Biol.* **214**, 699–710 (1990).
- White, H. D., Belknap, B. & Jiang, W. J. *J. biol. Chem.* **268**, 10039–10045 (1993).
- Shimizu, T. et al. *J. Cell Biol.* **112**, 1189–1197 (1991).
- Sellers, J. R., Eisenberg, E. & Adelstein, R. S. *J. biol. Chem.* **257**, 13880–13883 (1982).
- Warshaw, H. M., Derosiers, J. M., Work, S. S. & Trybus, K. M. *J. Cell Biol.* **111**, 453–463 (1990).
- Warshaw, D., Peterson, J. & Alpert, N. *Biophys. J.* **59**, 186a (1991).
- Sussman, M. in *Dictyostelium Discoideum: Molecular Approaches to Cell Biology* (ed. Spudich, J. A.) 9–29 (Academic, Orlando, 1987).

ACKNOWLEDGEMENTS. We thank H. Warrick for reviewing the manuscript. This work was supported by grants from the NIH to J.A.S. T.Q.P.U. was supported by a fellowship from the American Heart Association, California affiliate. K.M.R. is a trainee of the Medical Scientist Training Program at the Stanford University School of Medicine.

CORRECTION

Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair

Micheline Strand, Tomas A. Prolla, R. Michael Liskay & Thomas D. Petes

Nature **365**, 274–276 (1993)

THE penultimate sentence in the legend to Table 1 of this letter should read: "The plasmid pSH91 contains an in-frame insertion of poly(GT) (33 bp) in *URA3*; this plasmid is identical to pSH44 (ref. 8) except for the orientation of the tract", and not that the plasmid is identical to pSH36 apart from the size of the tract, as originally stated. This error does not affect our conclusions. □

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list: grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.

Shape-shifting serpins

Recent structures of members of the serine proteinase inhibitor families (serpins) continue to emphasize the flexibility of the reactive loop and suggest possible 'active' conformations.

PROTEINASES are essential for the life cycles of all organisms, carrying out vital functions such as precursor processing, digestion, defence and degradation. But if such molecules were given a free rein they would perpetrate all kinds of unwanted damage and destruction. So it is not surprising that for most proteinases there is a proteinase inhibitor that controls the activity of these vital, and potentially deadly, enzymes.

The ubiquitous serine-proteinase inhibitors (serpins for short) constitute a family of such inhibitors: one of the most abundant proteins in beer — claimed to be responsible, in part, for its taste — is a barley serpin and one in ten proteins in our blood plasma is a serpin. Serpins regulate proteinases involved in a range of key biological processes that include blood coagulation, fertilization and viral inhibition of inflammation; indeed it is not surprising that serpin dysfunction is associated with lung, liver and blood coagulation diseases such as emphysema, liver cirrhosis, thrombosis and pulmonary embolism. Their mode of action is therefore of more than academic interest. Now two new serpin structures, a modified human antichymotrypsin reported in this month's *Nature Structural Biology*^{1,2}, and a non-covalent dimer of antithrombin III molecules^{3,4}, have moved us closer to a picture of the active conformation of serpins and an understanding of the features of the inhibition reaction.

Like other protein inhibitors of the serine proteinases⁵, the business end of a serpin is a short 15-residue peptide — the reactive loop — on the surface of the molecule and is principally responsible for interacting with the active site of the target serine proteinase. The serpin forms a stable covalent complex with the proteinase which can then be removed from the extracellular milieu by binding to specific cellular receptors. Alternatively, the cleaved and inactivated serpin may be slowly released from the complex.

The structures of inhibitory serpins reveal a radical transformation upon cleavage^{6,7}. The reactive loop on the surface is cut between residues P1 (on the N-terminal side of the loop) and P1' (on the C-terminal side of the cleaved loop). The P1 side of the cleaved reactive loop then inserts into a large five-stranded β -sheet—sheet A—to become strand s4A, forming a six-stranded sheet, while the residues on the P1' side contribute to the smaller β -sheet C. Once contiguous amino acids in the intact molecule, the P1 and P1' residues are now up to 70Å apart in the cleaved molecule.

Not all serpins are inhibitory and the structure of one such non-inhibitory serpin, ovalbumin, has provided a first glimpse of the intact peptide which is the analogue of the reactive loop of the inhibitory serpins^{6,7}. In this protein the loop forms a regular α -helix, held above the surface of the protein on two short 'stalks'. In this conformation it is not a suitable substrate for the active site of a serine proteinase. Why can't the ovalbumin loop take on an extended conformation that would bind serine proteinases? Isolated peptides of varying length bind to and insert in the A-sheet in the normal position of the P1–16 reactive loop segment in inhibitory serpins, indicating that partial or pre-insertion of residues up to P14 in the s4A strand is required for inhibitor function. A series of bulky residues in the N-terminal stalk of ovalbumin reduces the flexibility of this region and prevents reinsertion of the surface peptide back into sheet A.

The conformational versatility of this family of molecules is further demonstrated by the structure of the 'latent', inactive form of human plasminogen-activator inhibitor-1. Here the reactive loop is fully inserted into the A-sheet as strand s4A even though the molecule has not been cleaved^{6,7}. But until now what has remained tantalizingly elusive is the structure of the so-called 'active' conformation, the form that interacts with the target serine proteinases.

In the antithrombin dimer^{3,4} the reactive loop of one monomer is inserted into the C β -sheet of the other monomer, the latter having a cleaved/latent conformation. The reactive loop in the inserted monomer is in a conformation that is similar, but not identical, to that of other peptide-based proteinase inhibitors.

The N-terminal stalk is partially inserted up to residue P14 in the A-sheet

between strands s5a and s3a.

How closely this mimics the real serpin–proteinase complex remains to be seen. Certainly it provides a framework for thinking about the antithrombin–thrombin interaction. But some features are at odds with expectations, for example the P1 residue, arginine 393, is pointing back towards the parent molecule rather than out towards the target molecule, and heparin, a potent stimulatory cofactor for the interaction that binds antithrombin, is not included in the complex.

The structure from Wei and colleagues¹ provides a further twist in this tale of proteinase inhibitors. In this uncleaved and uncomplexed serpin, a variant of anti-chymotrypsin, a reactive loop is revealed with a very distorted helical structure. The active-site P1 residue, important in specificity and activity, is pointing out into solution, away from the serpin scaffold. As such, the primary recognition event could be docking with the proteinase S1 subsite.

Part of the N-terminal stalk, the equivalent of which is inserted in the active molecule in the antithrombin complex, is not pre-inserted, although there is an inviting 5Å gap at the 'top' of strands s5A and a3A. Wei *et al.* speculate that as the interaction proceeds, the reactive loop uncoils so that part of the stalk can then be inserted into the A-sheet, an idea consistent with immunological studies suggesting that insertion occurs after the serpin–proteinase complex forms. At some point, the active-site covalent link locks the two molecules together. When the serpin is subsequently released it will adopt the inactive conformation.

There is still a way to go. Further details of the reaction coordinate need to be defined, as do the roles of the various cofactors in the interaction.

Guy Riddihough

Guy Riddihough is Editor of Nature Structural Biology

Also in this month's *Nature Structural Biology*: critical residues for SH3 domain/peptide interactions; new ligand-binding pathways in myoglobin; haem doming as the primary event in haemoglobin ligand binding; structural dynamics in an electron-transfer complex; calbindin D_{9k} in the absence of Ca²⁺; dendrotoxin-I structure; *ab initio* prediction of a lysozyme-antibody complex; and are there knots in proteins?

1. Wei, A., Rubin, H., Cooperman, B.S. & Christianson, D.W. *Nature struct. Biol.* **1**, 250–257 (1994).
2. Fletterick, R.J. & McGrath, M.E. *Nature struct. Biol.* **1**, 201–203 (1994).
3. Schreuder, H.A. *et al.* *Nature struct. Biol.* **1**, 48–54 (1994).
4. Carrell, R.W., Stein, P.E., Fermi, G. & Wardell, M.R. *Structure* (in the press).
5. Bode, W. & Huber, R. *Curr. Opin. struct. Biol.* **2**, 45–52 (1991).
6. Carrell, R.W. & Evans, D.L.I. *Curr. Opin. struct. Biol.* **2**, 438–446 (1992).
7. Engh, R.A., Schulze, A.J., Huber, R. & Bode, W. *Behring Inst. Mitt.* **93**, 41–62 (1993).